Parametric Coding of Spatial Audio

Ph.D. Thesis

Christof Faller, July 9, 2004

Thesis advisor: Prof. Martin Vetterli

Audiovisual Communications Laboratory, EPFL Lausanne

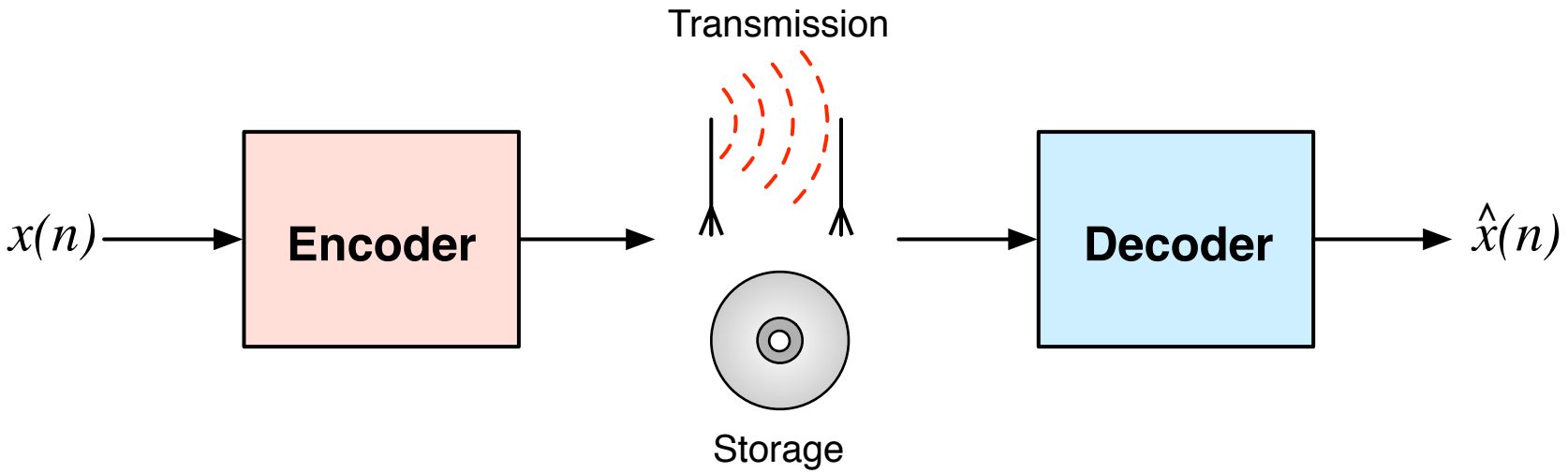
Parametric Coding of Spatial Audio

Contents:

- Audio Coding and Thesis Motivation
- Background
- Binaural Cue Coding (BCC)
- Variations of BCC
- Source Localization in Complex Listening Scenarios
- Conclusions

Audio Coding

Audio coding



Convert audio signal into a representation suitable for:

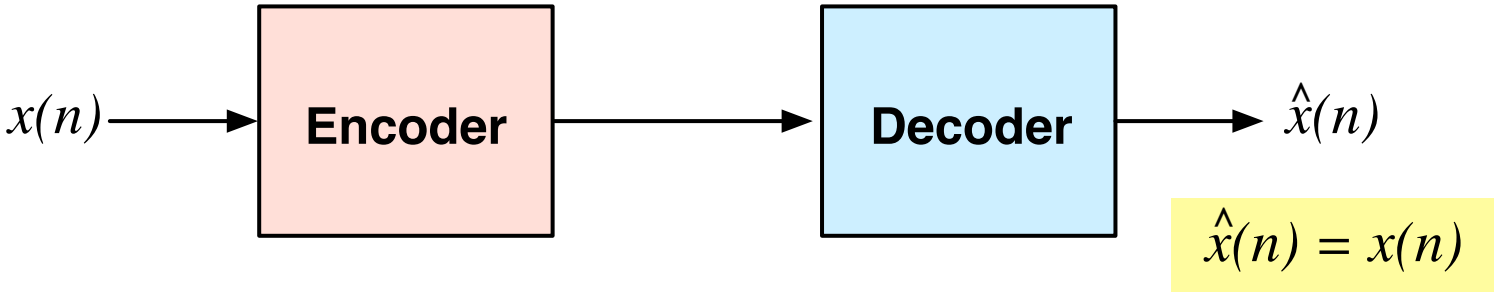
- Transmission
- Storage

Minimize bitrate

- Optimal for storage
- Optimal for transmission (source/channel coding)

Audio Coding

Lossless coding



Redundancy reduction

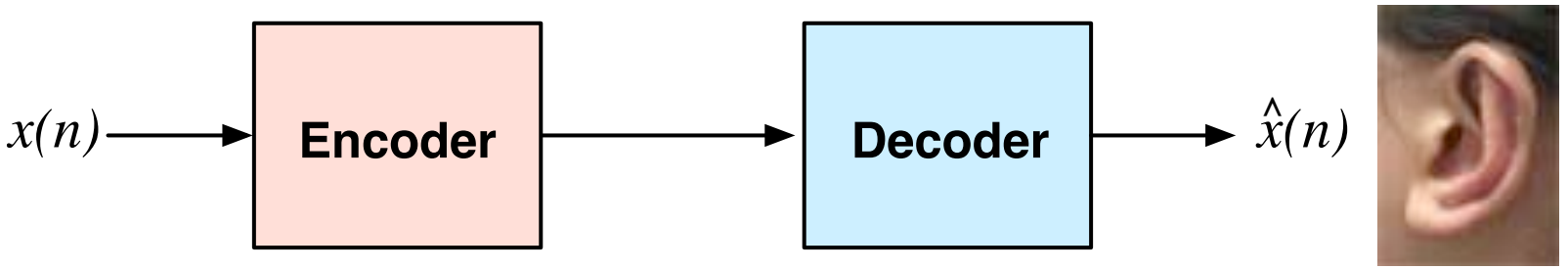
Bitrate example

CD stereo signals: $2x16x44100 = 1411$ kb/s

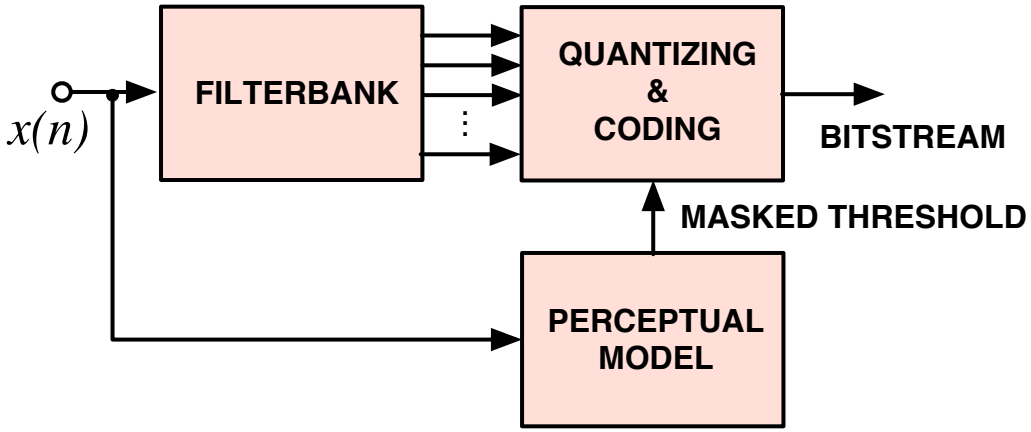
Lossless coding: ≈ 700 kb/s

Audio Coding

Perceptual audio coding



Redundancy reduction and receiver model

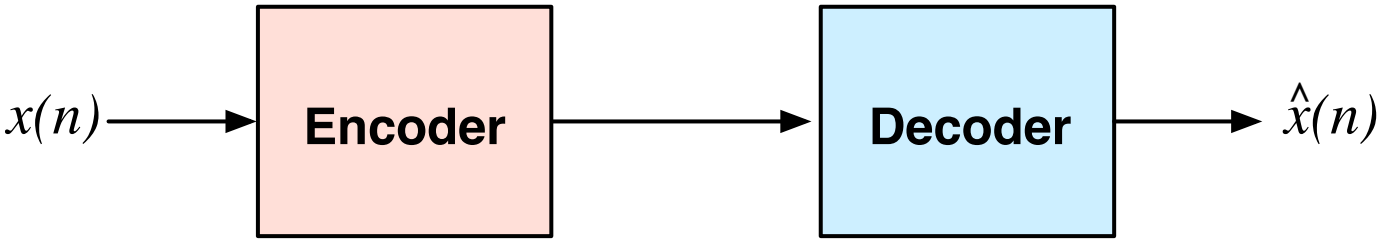


Bitrate example

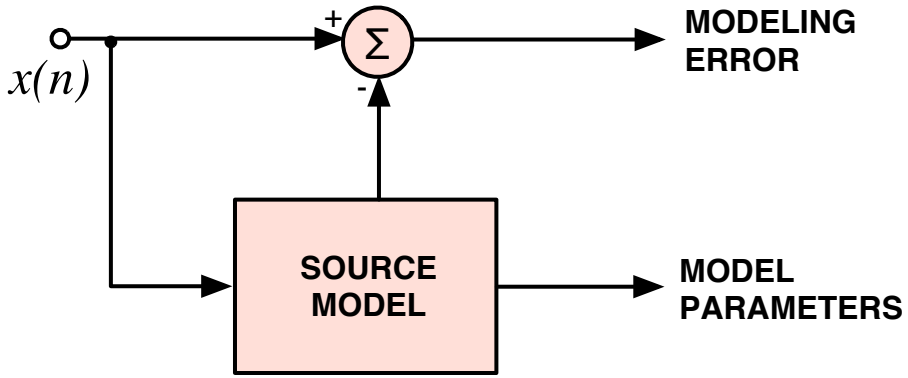
Stereo CD: *1411* kb/s, Perceptual Audio Coding: ≈ *140* kb/s

Audio Coding

Parametric audio coding



Redundancy reduction and source model



Bitrate example

4-32 kb/s

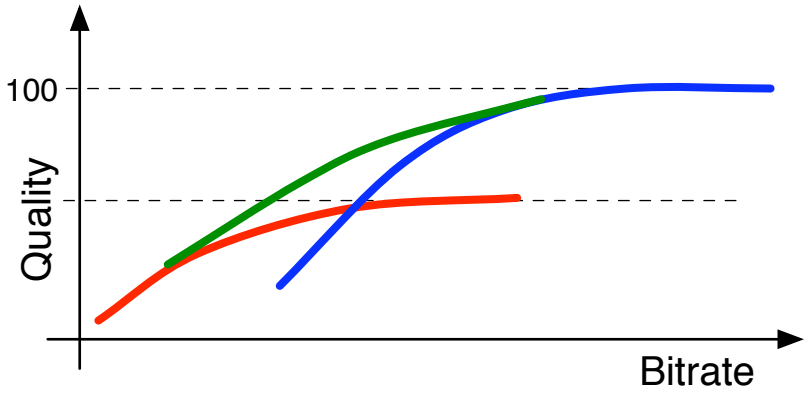
Audio Coding

Lossless audio coding:
- no quality loss
- low compression ratio

Perceptual audio coding:
- (ideally) no quality loss
- medium compression ratio

Parametric audio coding:
- quality loss
- high compression ratio

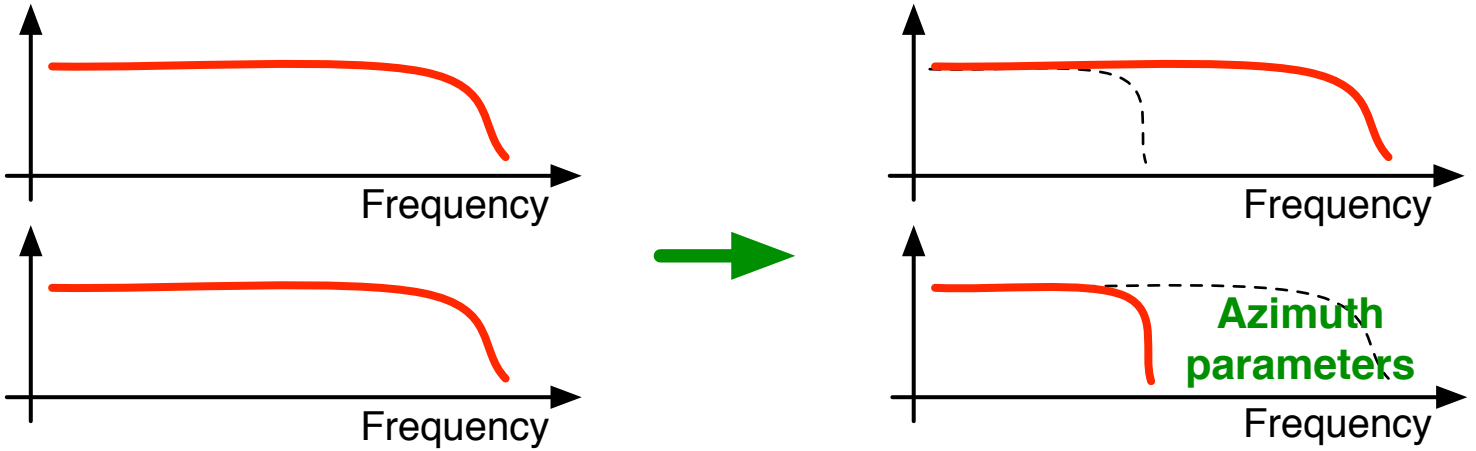
Hybrid audio coding:
- compromise between quality loss and bitrate



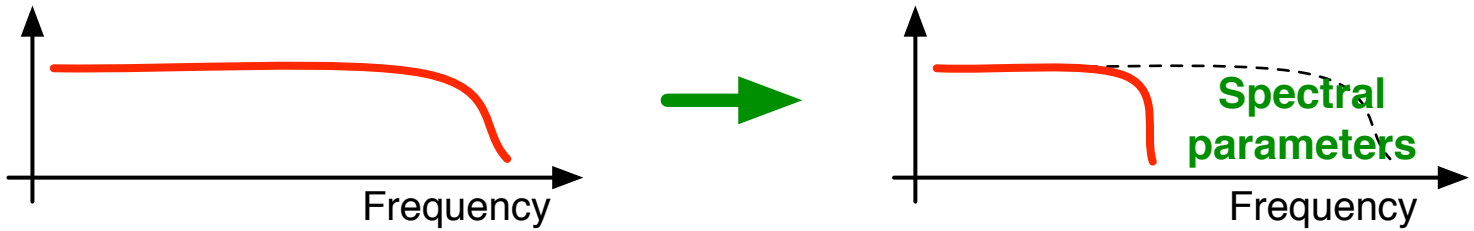
Audio Coding

Hybrid audio coding:

Intensity stereo coding (ISC):



Spectral bandwidth replication (SBR):



Thesis Motivation

Audio coding state of the art (2000):

Source properties:
- -vocal tract / instrument body properties
- -harmonicity
- -sinusoids/transients/noise

Receiver properties:
- -masking
- -binaural masking
- -frequency selectivity
- -importance of level difference cues at higher frequencies

Thesis Motivation

Audio coding state of the art (2000):

Source properties:
- -vocal tract / instrument body properties
- -harmonicity
- -sinusoids/transients/noise
- **-spatial recording/mixing model**

Receiver properties:
- -masking
- -binaural masking
- -frequency selectivity
- -importance of level difference cues at higher frequencies
- **-perception of auditory spatial image**

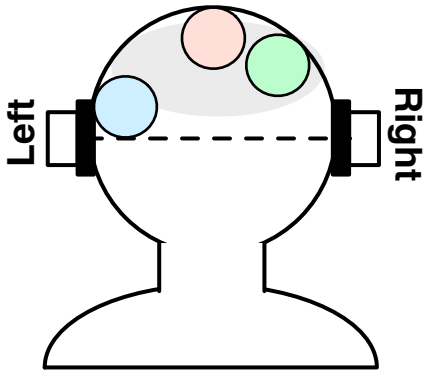
Parametric Coding of Spatial Audio

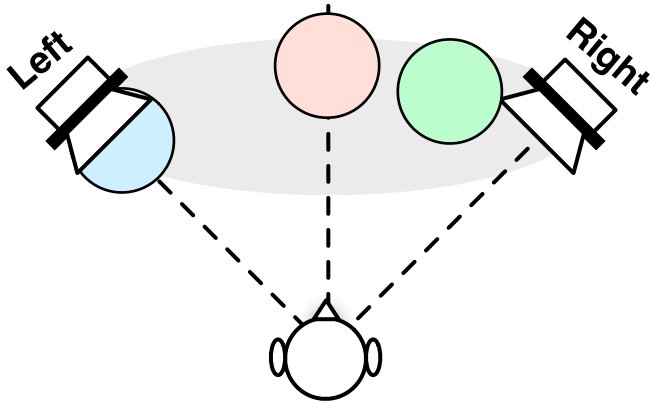
Contents:

Background

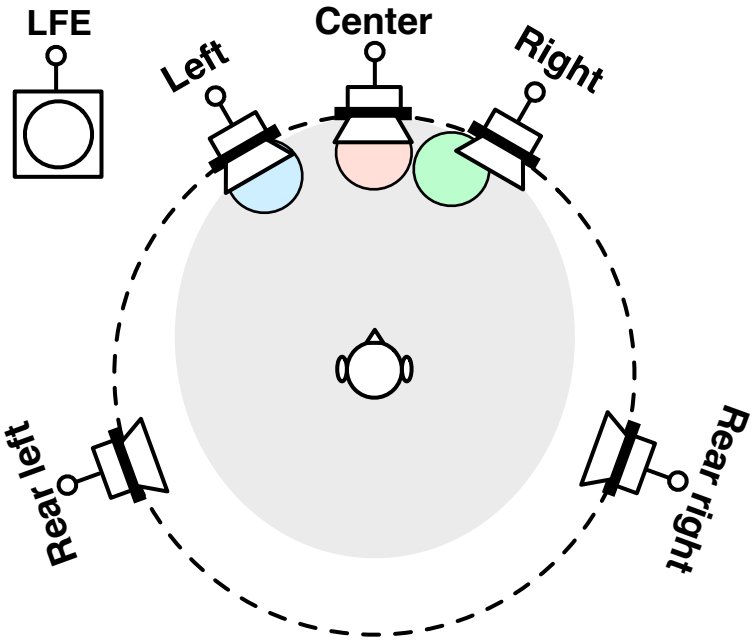
Spatial audio playback

Two-channel stereo
(headphone and loudspeaker playback)





5.1 Surround

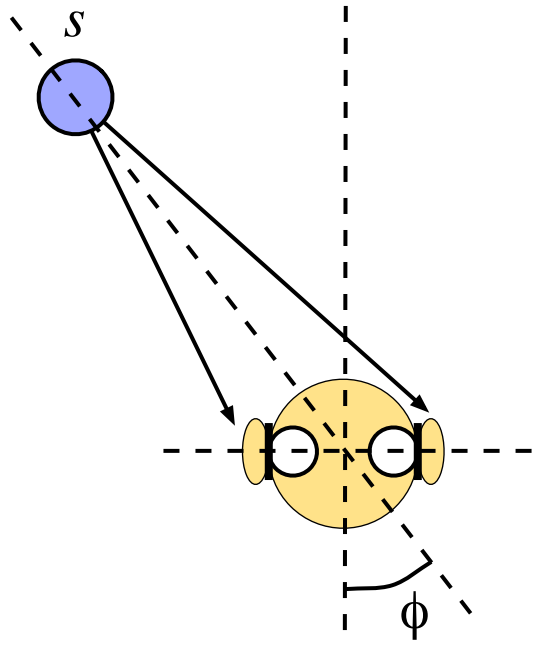


Spatial audio playback:
- perception of an auditory spatial image

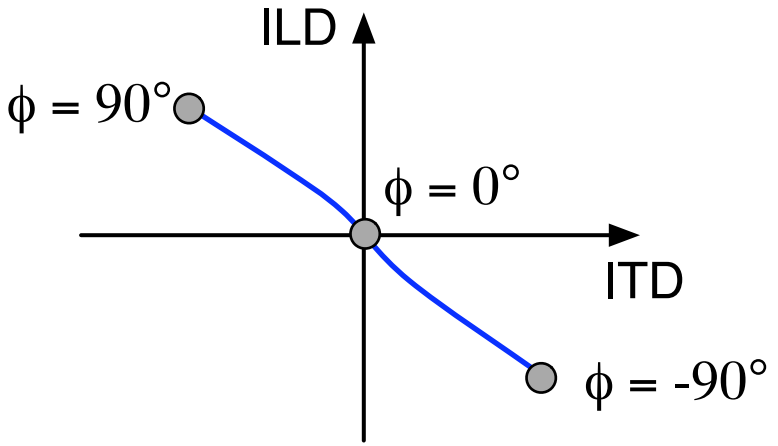
Background

Interaural differences

One source, free-field



Source azimuth and ITD/ILD:
(Narrowband signal)



→ distance difference,
head shadowing

→ interaural time and level difference
(**ITD and ILD**)

Background

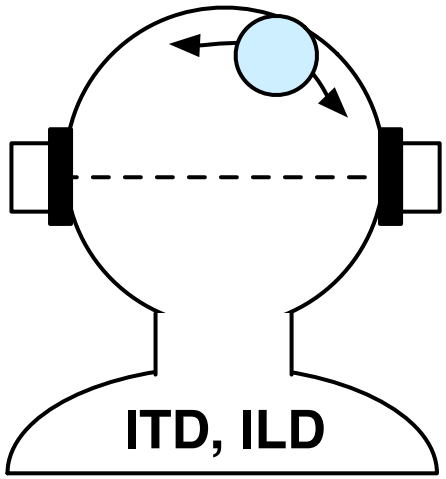
Interaural differences

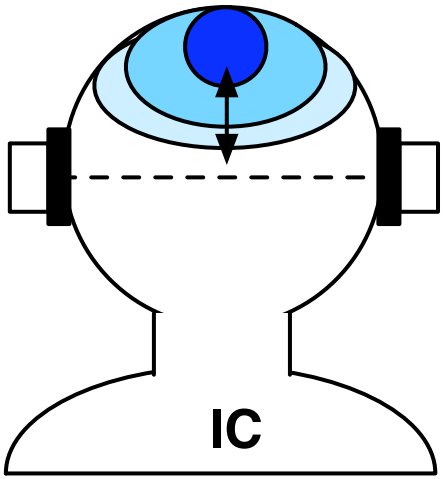
$$\mathrm{ITD}(n) = \arg\max_d \{\Phi(d,n)\} \quad \Phi(d,n) = \frac{E\{e_1(n)e_2(n-d)\}}{\sqrt{E\{e_1^2(n)\}E\{e_2^2(n-d)\}}}$$

$$\mathrm{ILD}(n) = 10\log_{10}\left(\frac{E\{e_2^2(n)\}}{E\{e_1^2(n)\}}\right)$$

$$\mathrm{IC}(n) = \max_d |\Phi(d,n)|$$ (interaural coherence)

Headphone playback





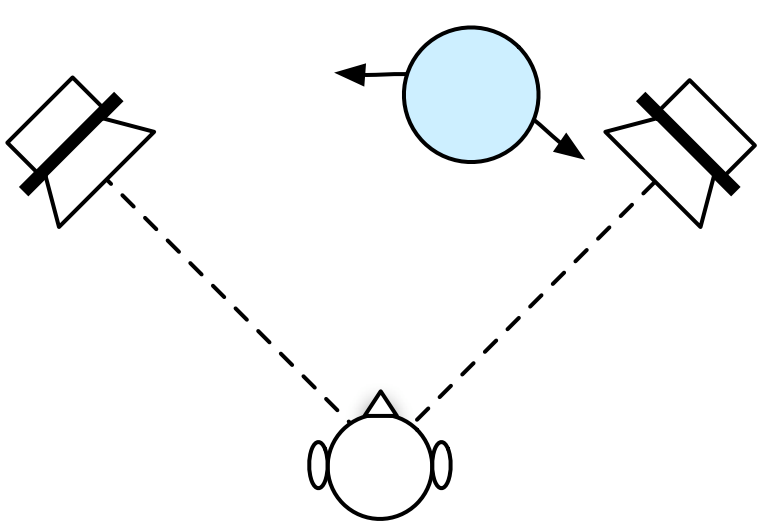
Background

Inter-channel differences:

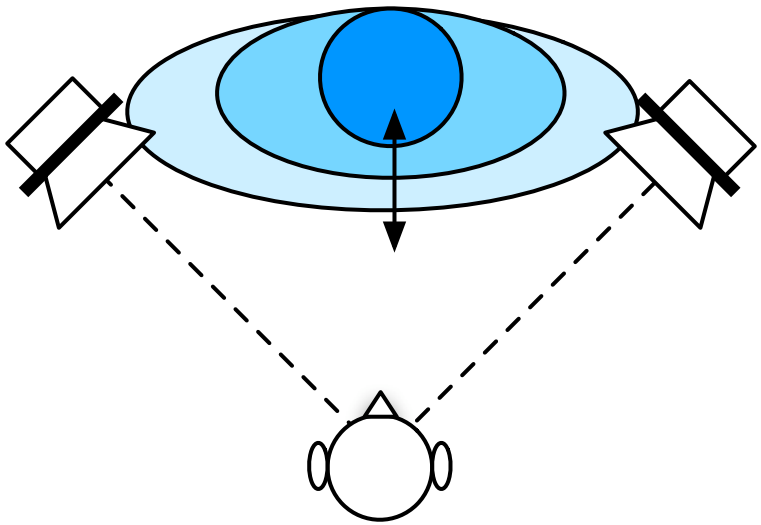
Inter-channel time-difference (**ICTD**)
Inter-channel level-difference (**ICLD**)
Inter-channel coherence (**ICC**)

Headphone playback: Inter-channel and interaural differences are the same

Loudspeaker playback



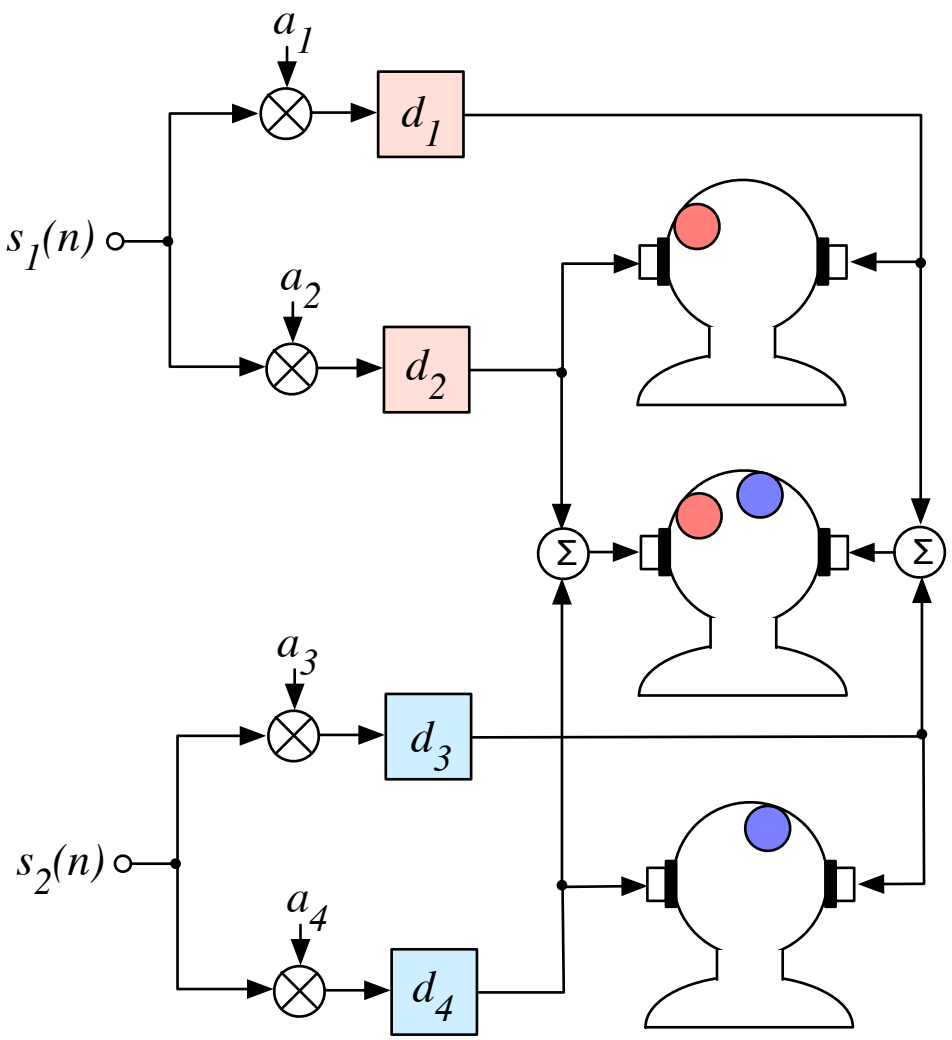
ICTD, ICLD

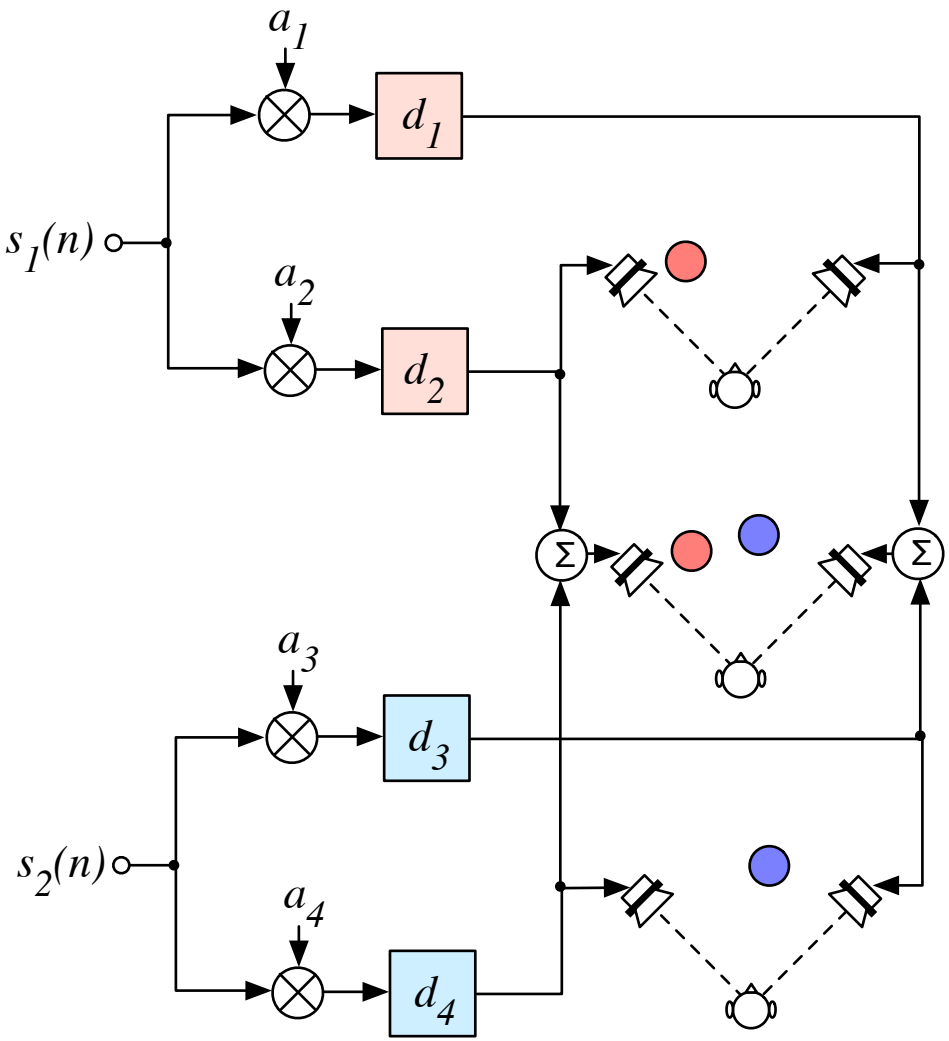


ICC

Background

Mixing stereo signals:

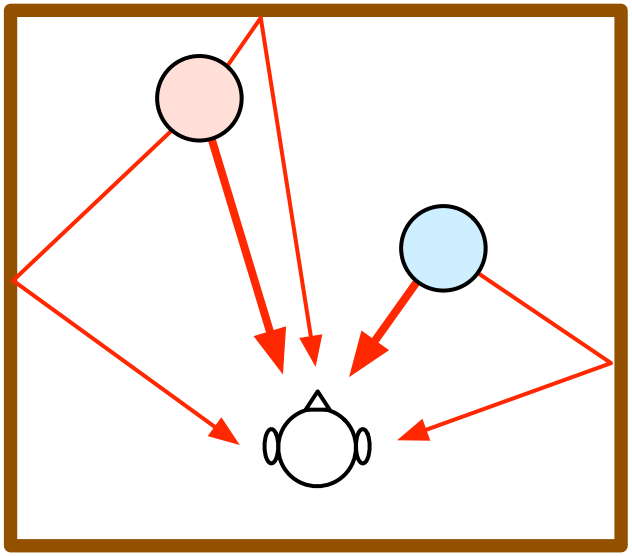




Background

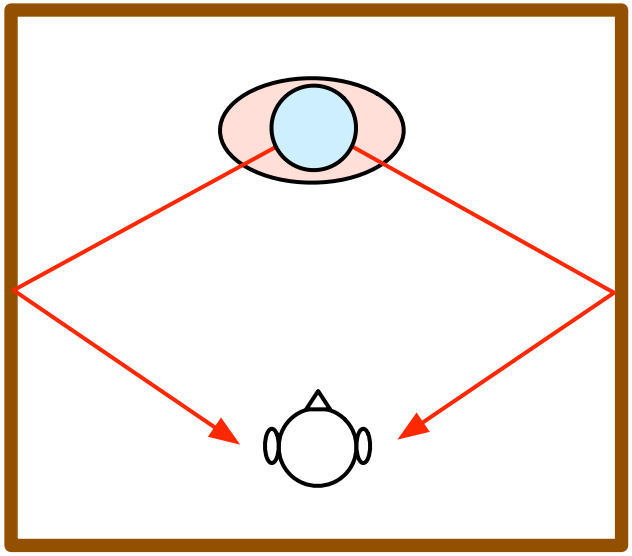
Other auditory spatial image attributes:

Auditory event distance



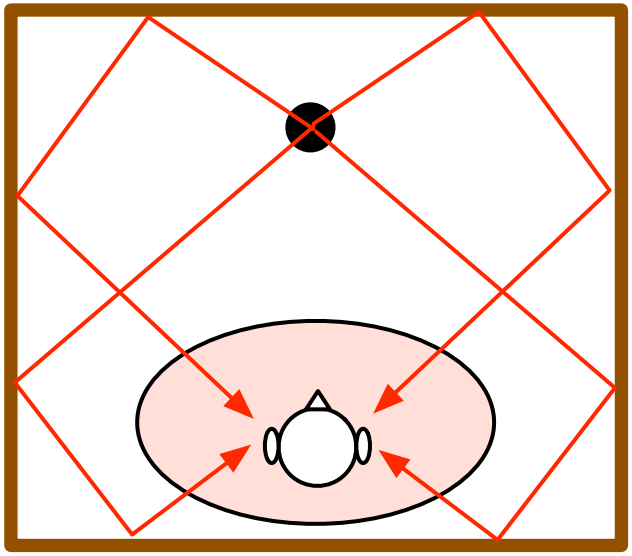
- Power of ear-input signals
- Ratio of power of direct to reflected sound

Auditory event width



- Lateral fraction

Listener envelopment



- Late lateral energy fraction

Spatial audio playback:

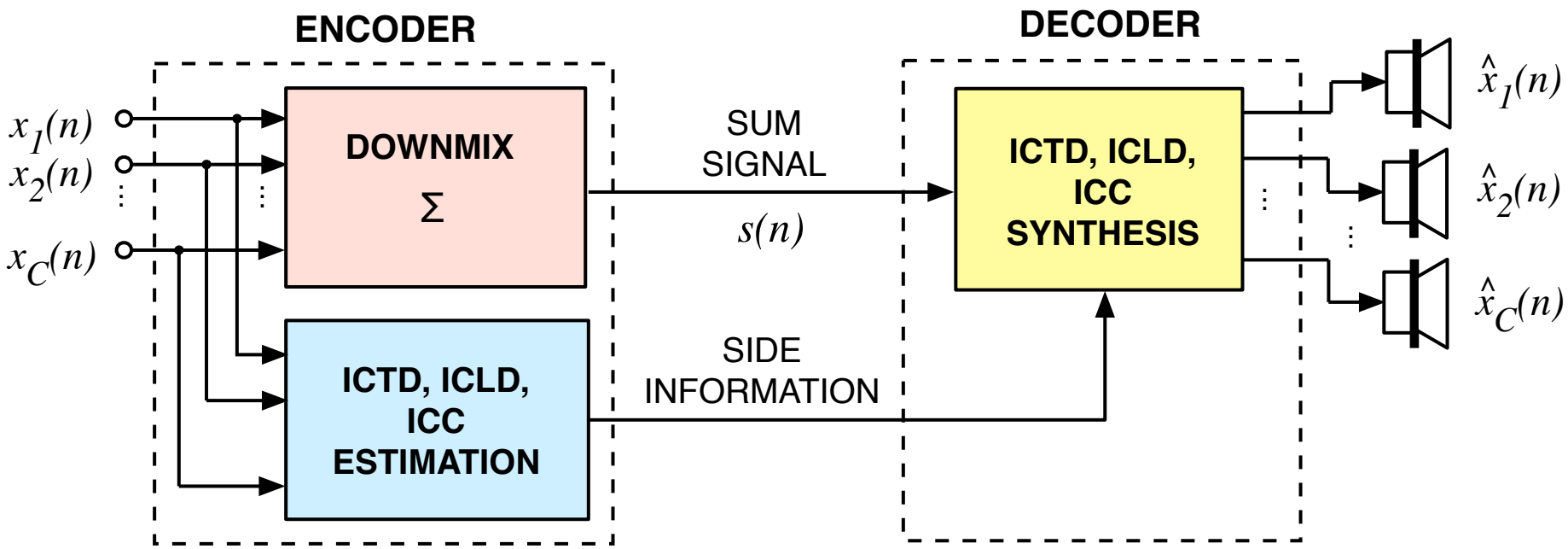
These attributes are controlled by adding reflections to the signal channels.

Parametric Coding of Spatial Audio

Contents:

- Audio Coding and Thesis Motivation
- Background
- **Binaural Cue Coding (BCC)**
- Variations of BCC
- Source Localization in Complex Listening Scenarios
- Conclusions

Binaural Cue Coding (BCC)

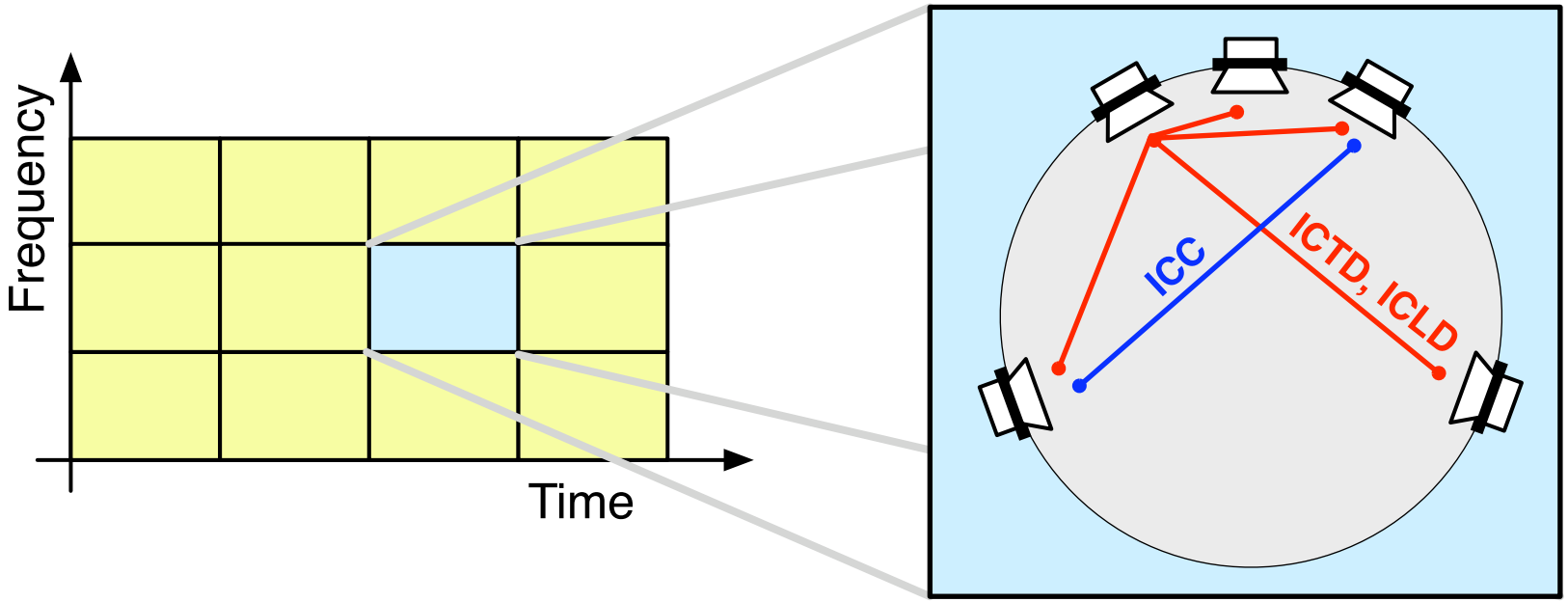


- N-to-1 channel downmix
- Estimation of inter-channel cues

- 1-to-N channel synthesis by means of inter-channel cue synthesis

Binaural Cue Coding (BCC)

Estimation of inter-channel cues: ICTD, ICLD, and ICC



ICTD/ICLD:

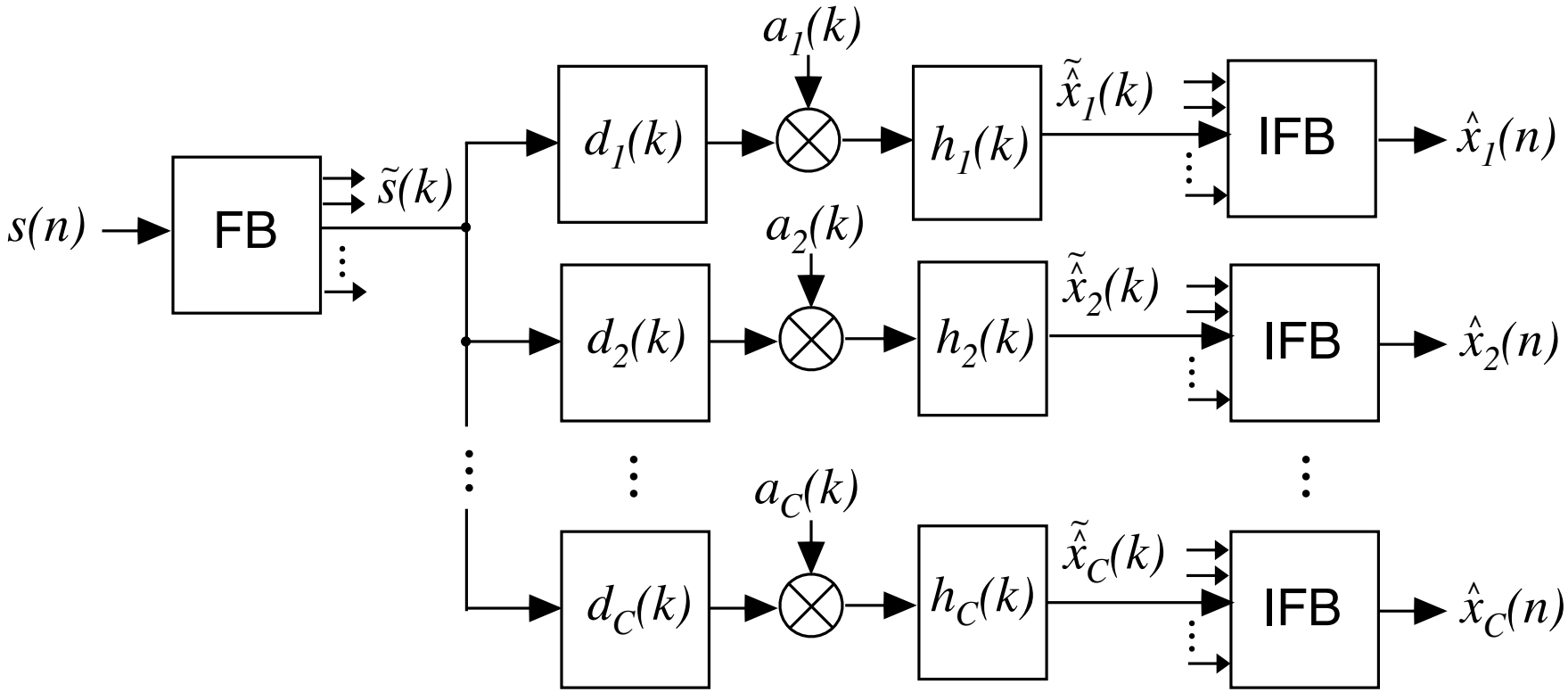
1-3 kb/s (per channel pair)

ICC:

1-3 kb/s

Binaural Cue Coding (BCC)

ICTD/ICLD/ICC synthesis

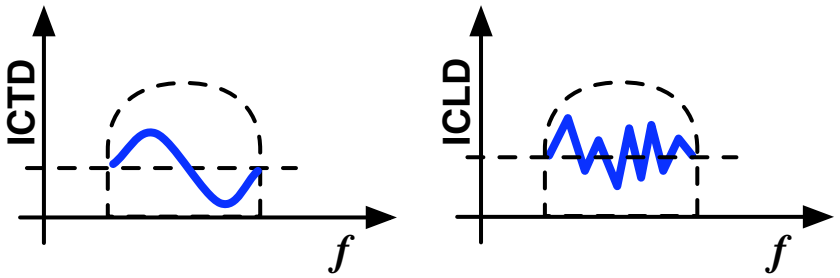


d_i : delays

a_i : scale factors

h_i : filters

h_i :

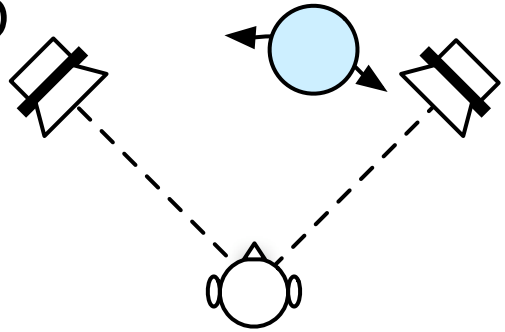


Binaural Cue Coding (BCC)

ICTD, ICLD, and ICC and auditory spatial image attributes

Auditory event localization:

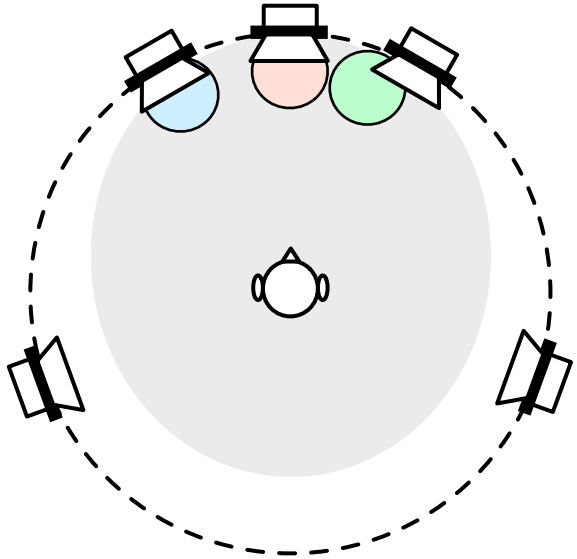
ICTD, ICLD



Effect of late reflections:
(ambience, listener envelopment,
auditory event distance)

ICC (de-correlation)

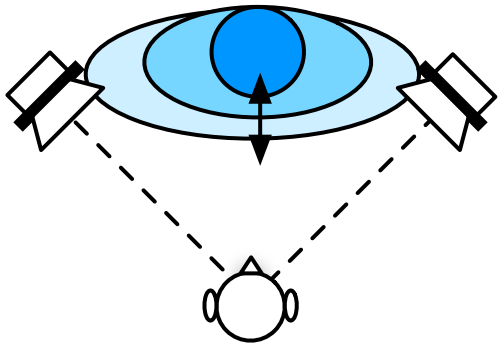
ICLD (reverberation decay)



Effect of early reflections:
(auditory event width, coloration)

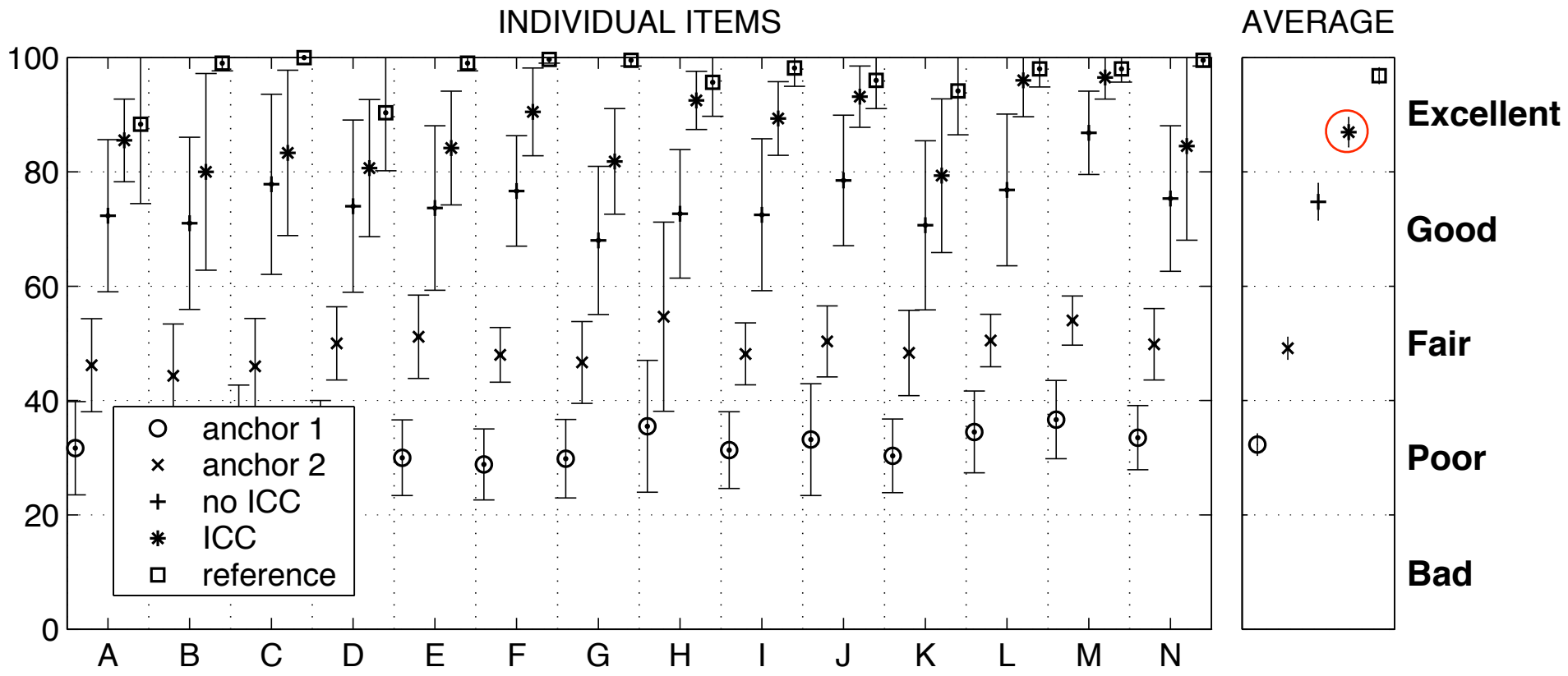
ICC (de-correlation)

ICLD (comb filter)



Binaural Cue Coding (BCC)

Subjective evaluation: Stereo audio quality



MUSHRA (ITU-R BS.1534), headphone listening, 7 experienced subjects

BCC: "excellent"

Binaural Cue Coding (BCC)

5.1 Demo:

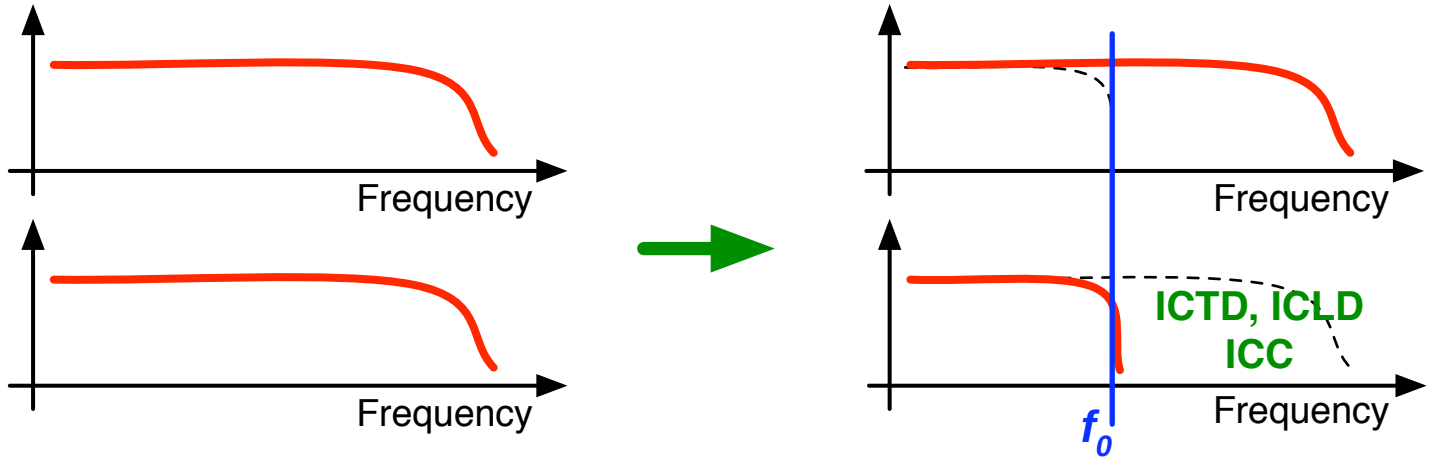
Reference
Sum signal
BCC + PCM
BCC + 32 kb/s coder (total 48 kb/s)

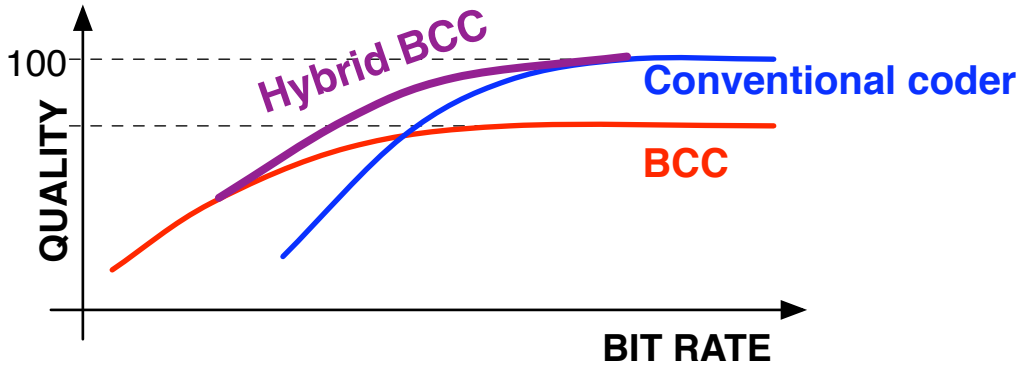
Parametric Coding of Spatial Audio

Contents:

Variations of BCC

Hybrid BCC



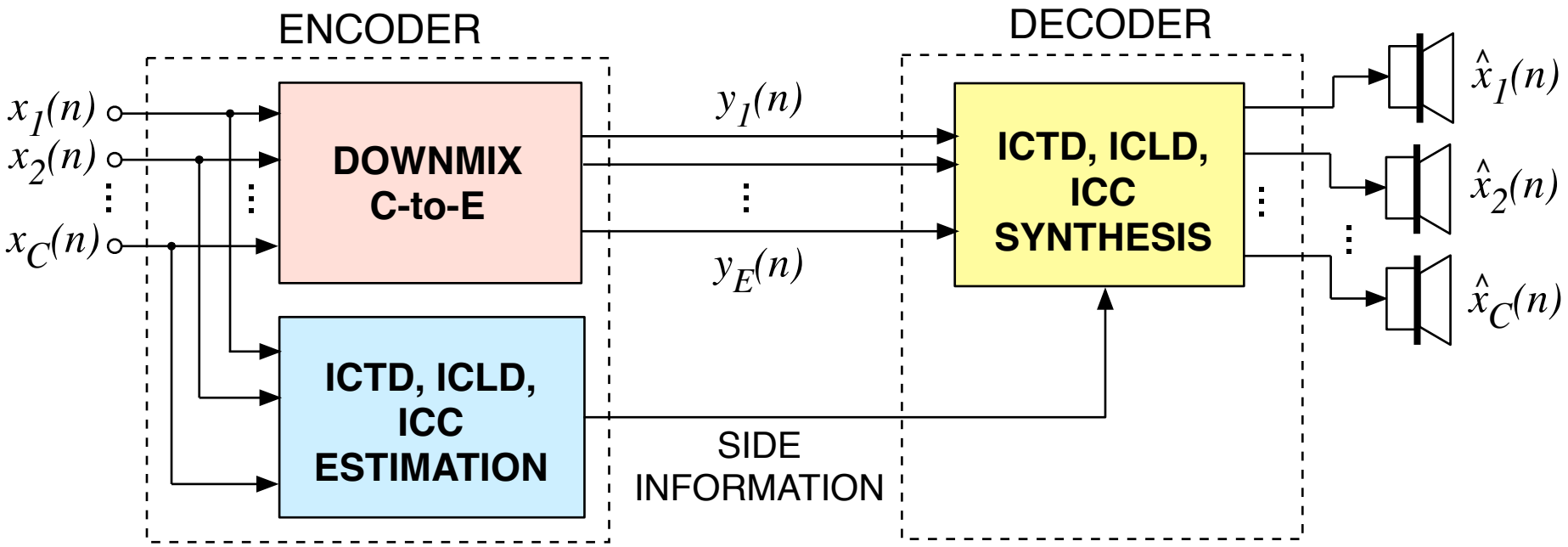


Differences to intensity stereo coding:

- consider more cues (not only intensities)
- use separate filterbank (better performance)

Variations of BCC

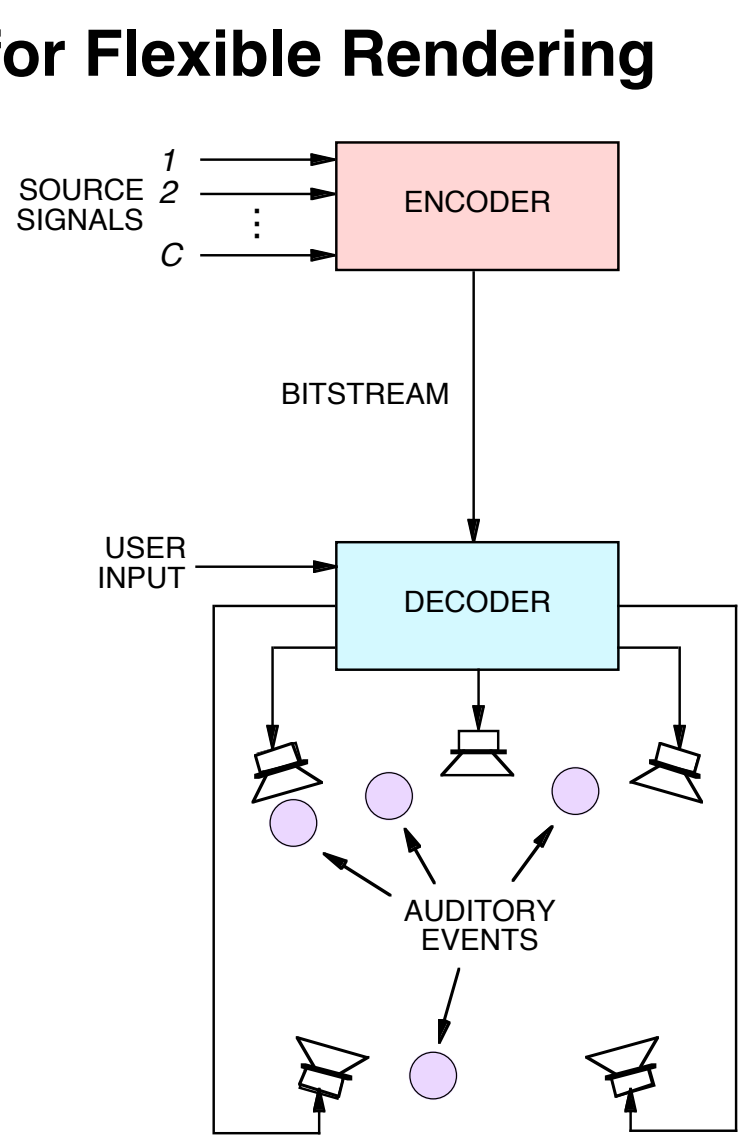
C-to-E BCC



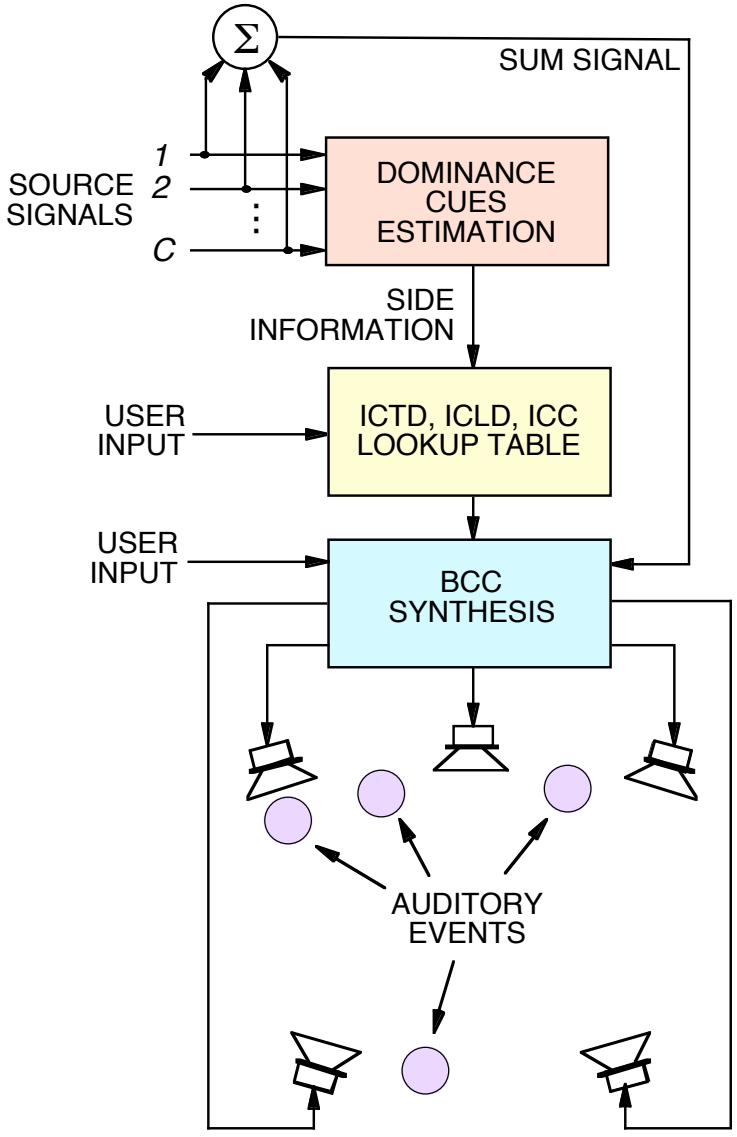
- Potentially better quality than C-to-1 BCC
- E-channel backwards compatible coding of C-channel audio

Variations of BCC

BCC for Flexible Rendering



Conventional system

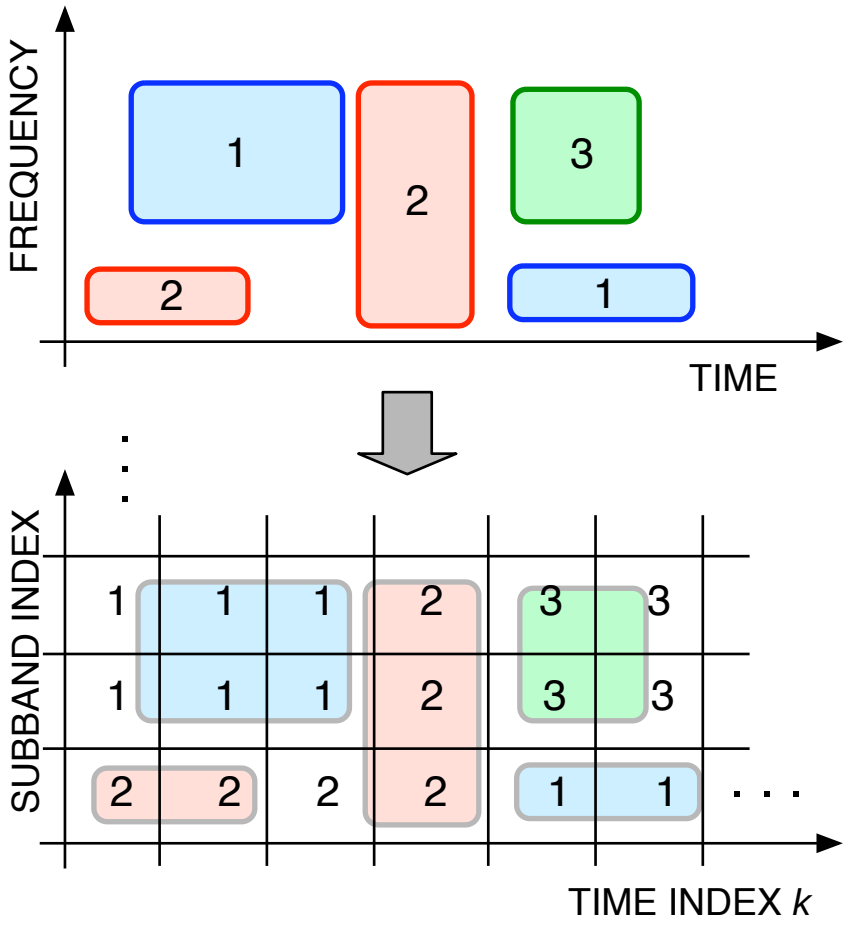


BCC for Flexible Rendering

Variations of BCC

BCC for Flexible Rendering

Side information:
Time-frequency structure of sum signal



Binaural Cue Coding (BCC)

BCC for Flexible Rendering Demo:

4 Instruments playing together:

Transmitted sum signal
BCC + PCM (2.5 kb/s BCC side information)

Parametric Coding of Spatial Audio

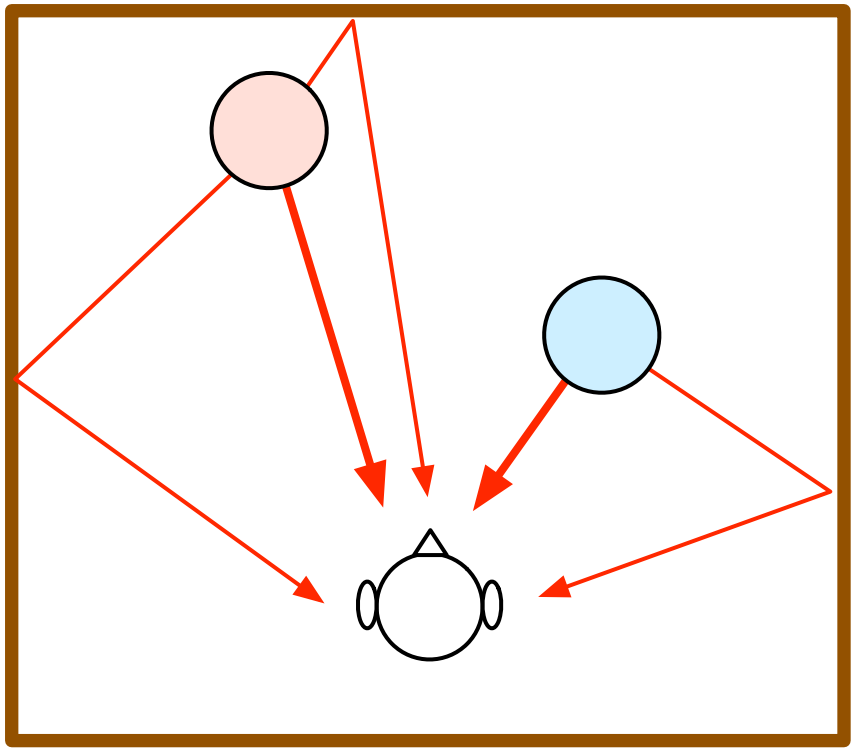
Contents:

- Audio Coding and Thesis Motivation
- Background
- Binaural Cue Coding (BCC)
- Variations of BCC
- **Source Localization in Complex Listening Scenarios**
- Conclusions

Source Localization in Complex Listening Scenarios

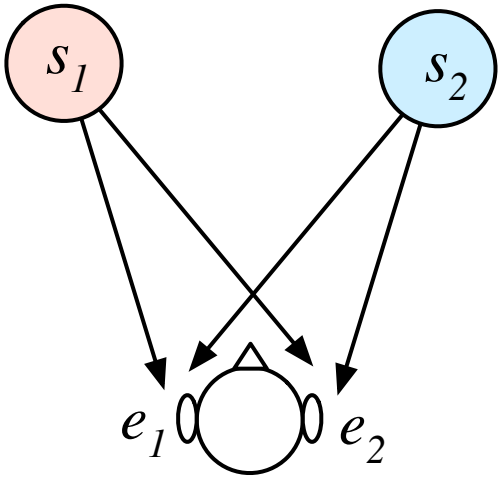
Model for source localization in complex listening scenarios:

- concurrently active sources
- direct and reflected sound

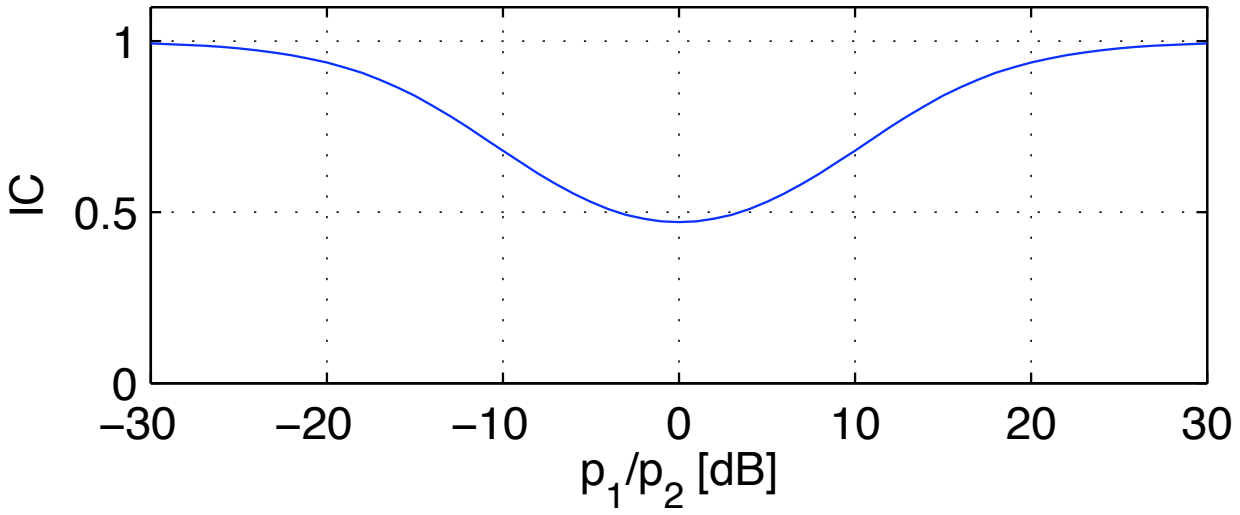


Source Localization in Complex Listening Scenarios

IC and concurrent sources



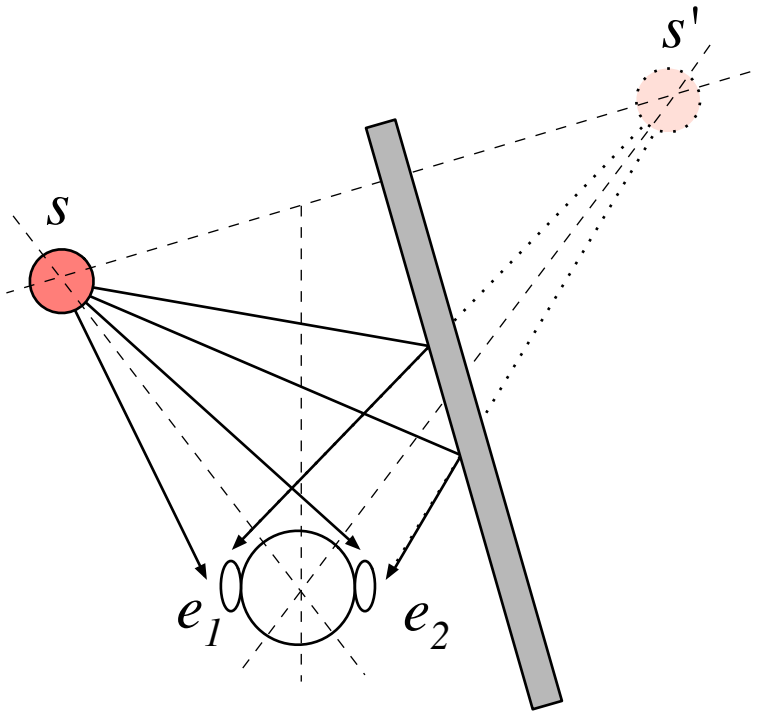
$$
\begin{aligned}
e_1(n) &= a_{11}s_1(n-d_{11}) + a_{21}s_2(n-d_{21}) \\
e_2(n) &= a_{12}s_1(n-d_{12}) + a_{22}s_2(n-d_{22}) \\
\mathrm{IC} &= \frac{\max\{a_{11}a_{12}p_1, a_{21}a_{22}p_2\}}{\sqrt{(a_{11}^2p_1 + a_{21}^2p_2)(a_{12}^2p_1 + a_{22}^2p_2)}}
\end{aligned}
$$

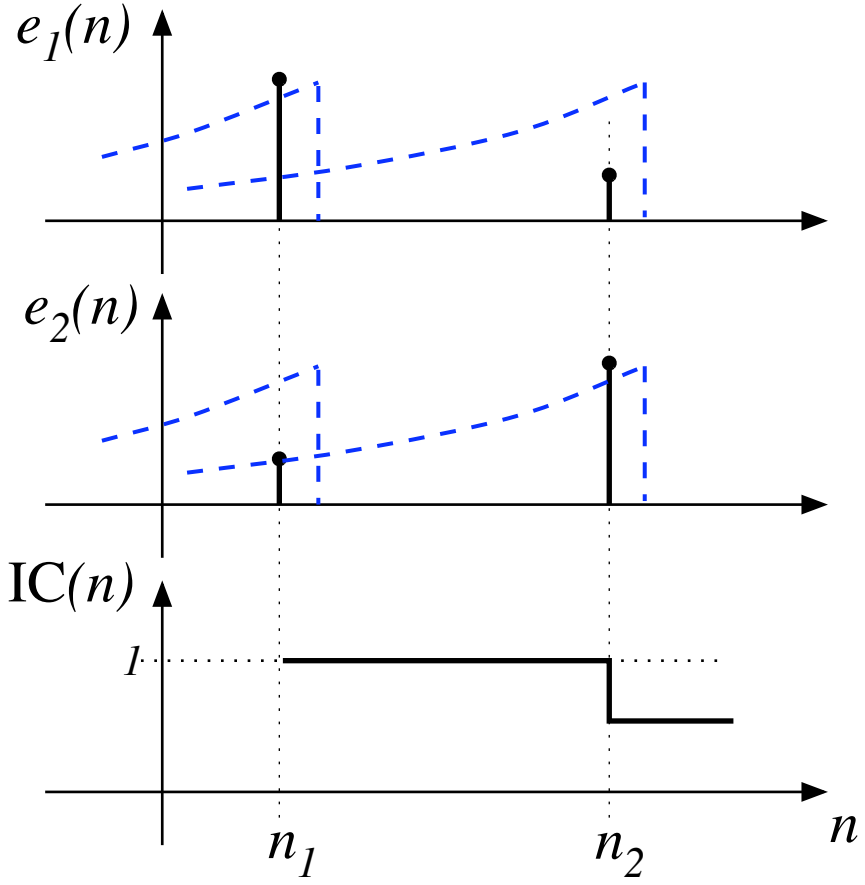


→ IC can be used as a **"single-source-active indicator"**

Source Localization in Complex Listening Scenarios

IC and reflections

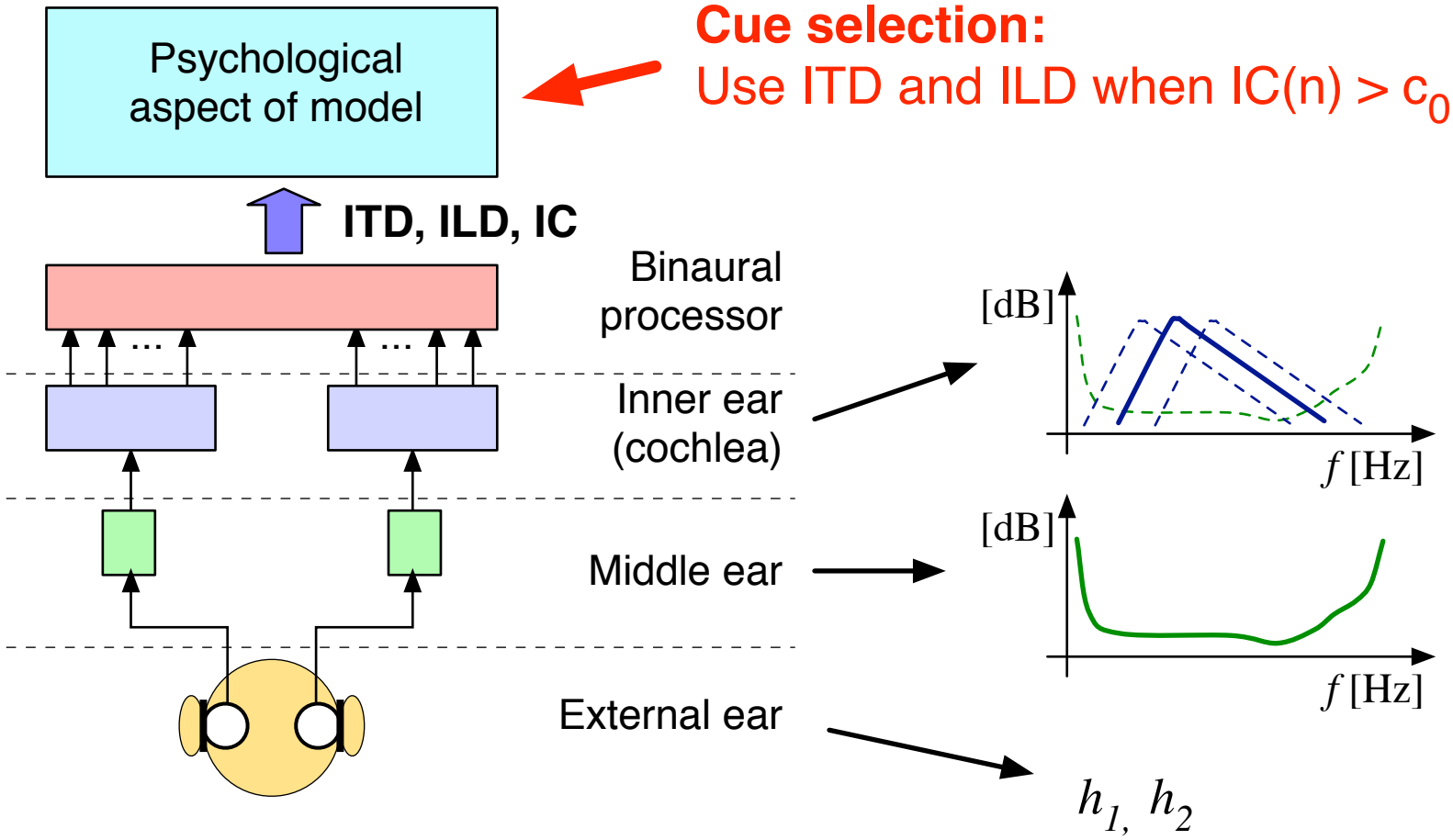




→ IC can be used as a **"first-wavefront indicator"**

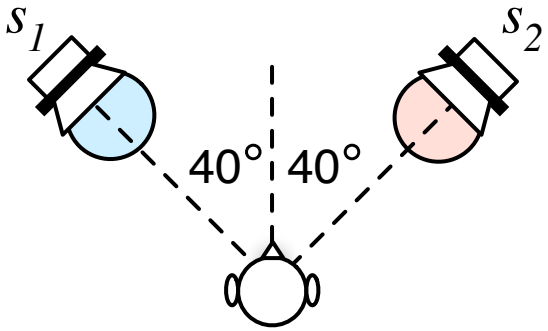
Source Localization in Complex Listening Scenarios

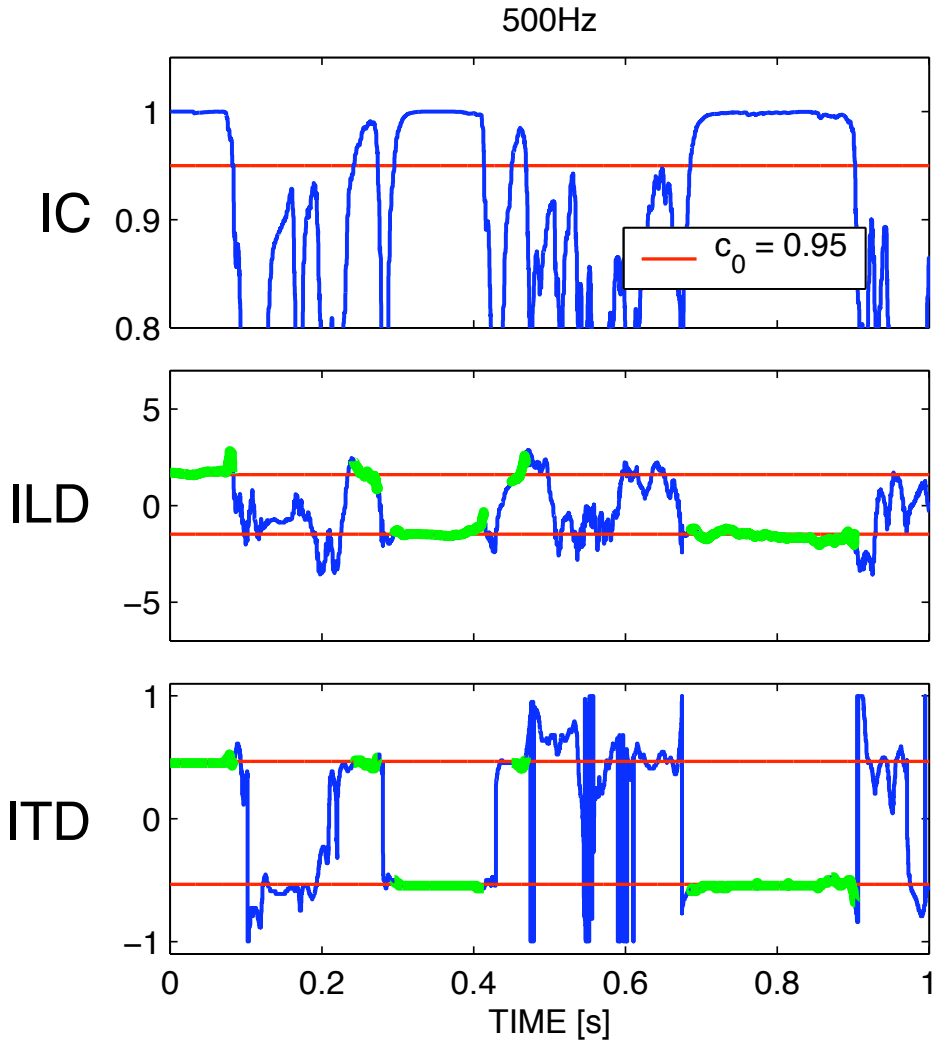
Model for source localization



Source Localization in Complex Listening Scenarios

Two concurrent male speech sources, free-field

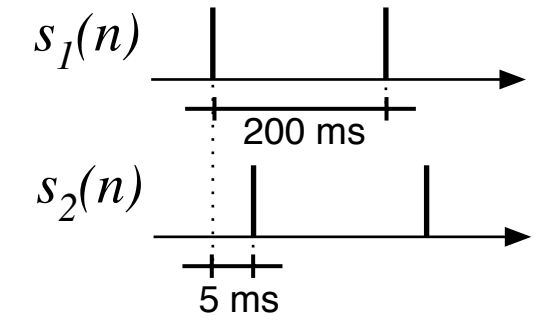


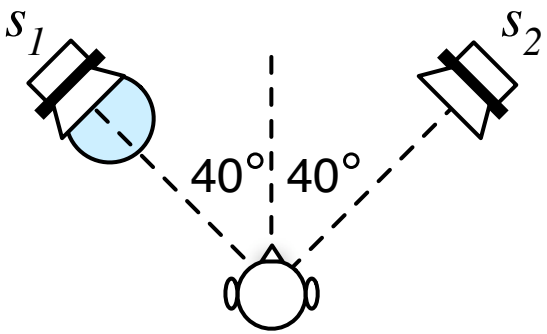


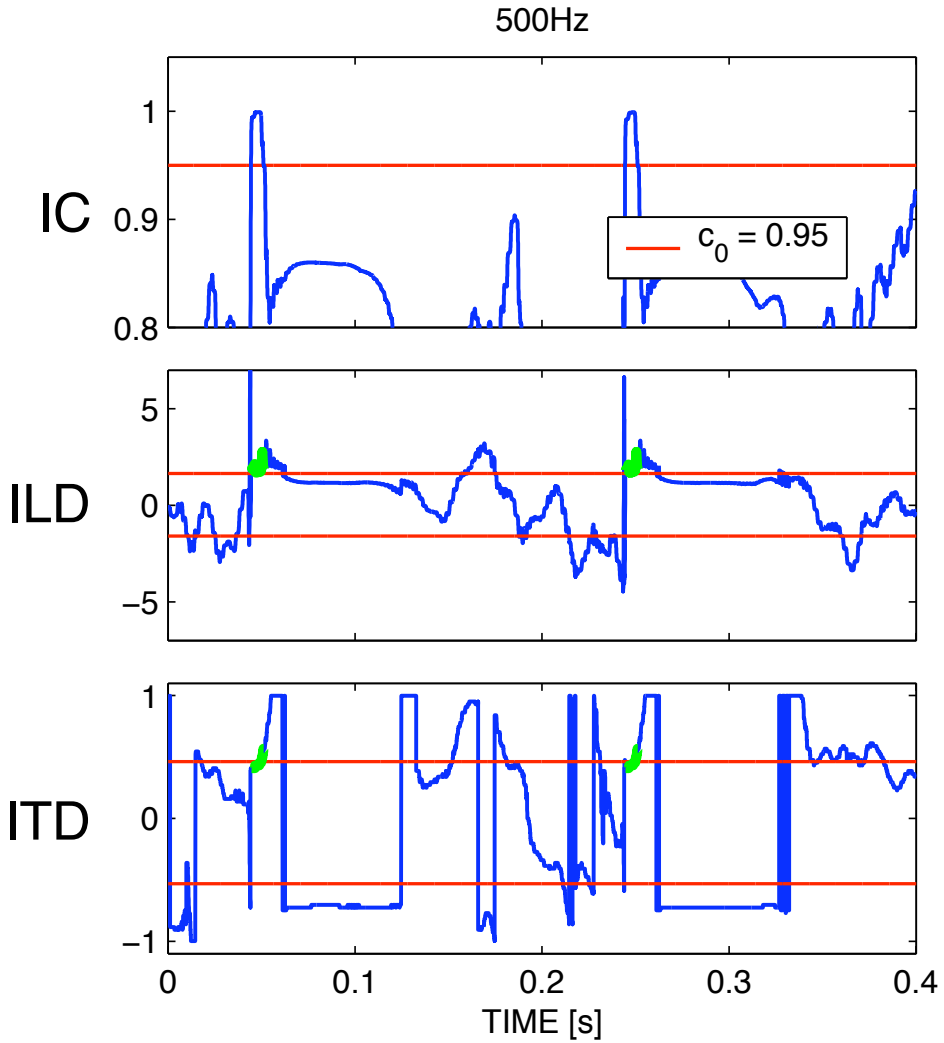
Source Localization in Complex Listening Scenarios

Precedence effect:

(Broadband pulses, 5ms lead/lag delay, free-field)

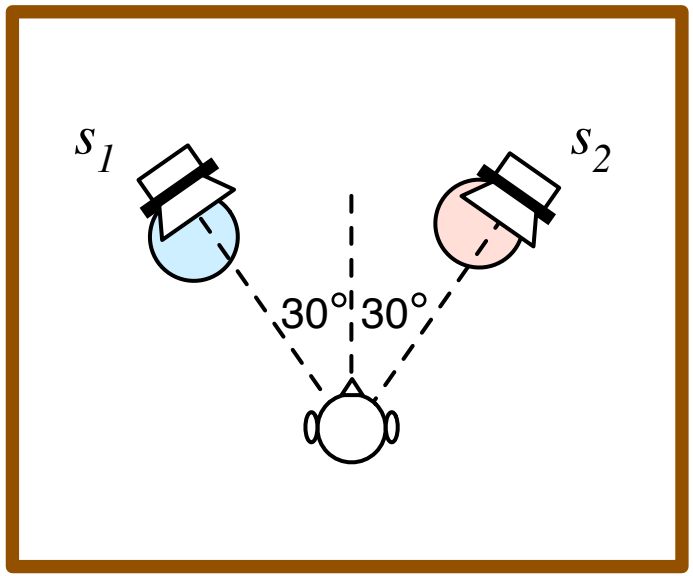


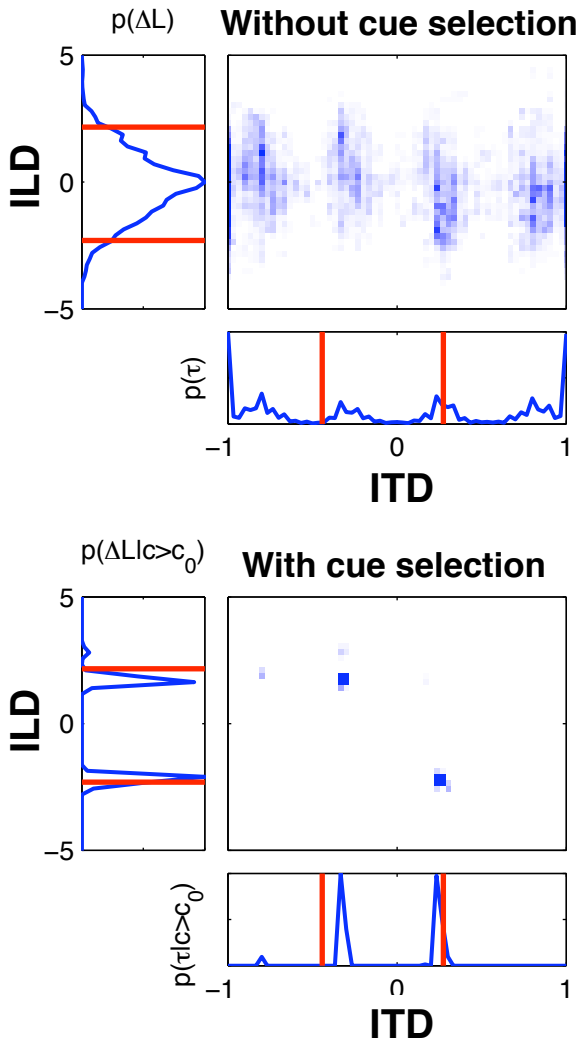




Source Localization in Complex Listening Scenarios

Source localization in reverberant environment:
(2 male speech sources, 500Hz: RT = 2s, 2kHz: RT = 1.4s)





Parametric Coding of Spatial Audio

Contents:

- Audio Coding and Thesis Motivation
- Background
- Binaural Cue Coding (BCC)
- Variations of BCC
- Source Localization in Complex Listening Scenarios
- **Conclusions**

Parametric Coding of Spatial Audio

Conclusions

Binaural Cue Coding (BCC):
- low bitrate coding of stereo and multi-channel audio signals
- low bitrate transmission of independent sources
- bridging between different audio formats

Proposed source localization model:
- speculates about role of IC for "cue selection"
- attempts at explaining source localization in complex listening

Parametric Coding of Spatial Audio

Future work:

Multi-channel BCC parameters: Different more efficient and flexible parametrization.

BCC applied to binaural recordings: Further investigate.

Psychoacoustic experiments with BCC stimuli.

Source localization model:

- Adaptation of cue selection threshold
- More simulations, psychoacoustic experiments
- Can model be related to other attributes of auditory spatial image?

Parametric Coding of Spatial Audio

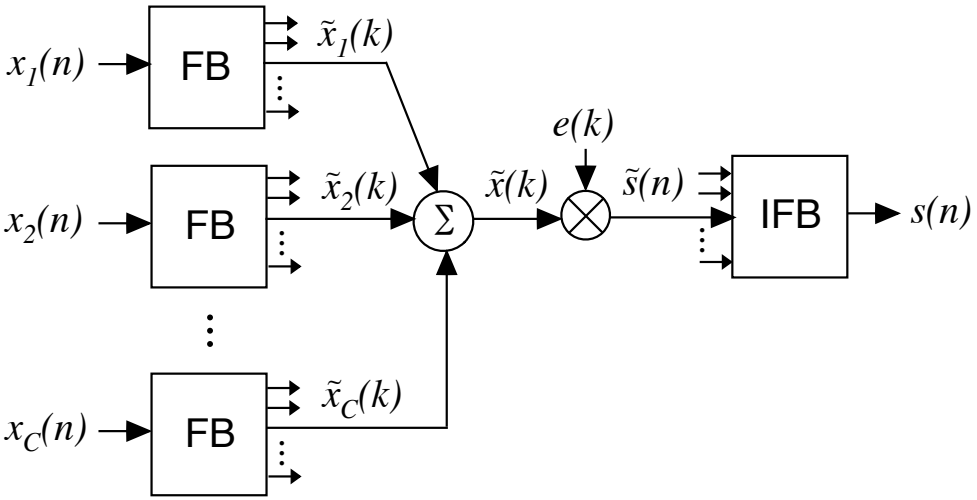
Acknowledgements:

Martin Vetterli, Peter Kroon, and all colleagues I collaborated with.

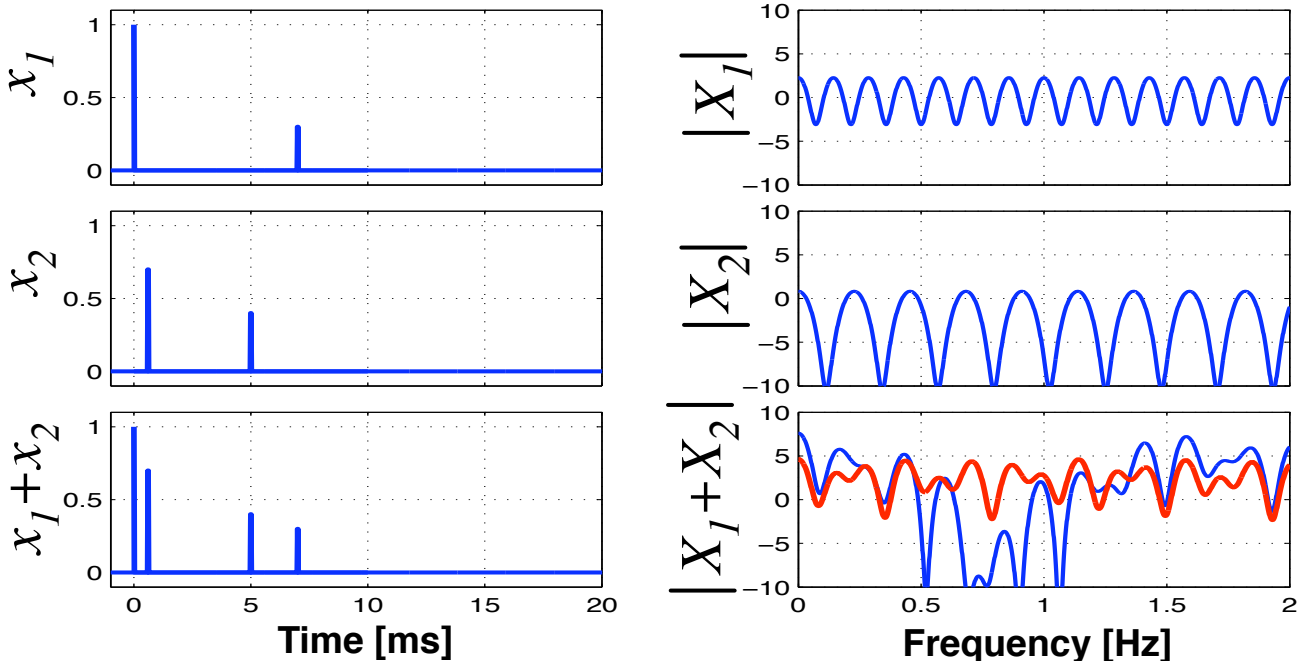
Thanks to all of you for coming to this defense!

Binaural Cue Coding (BCC)

Downmix with equalization

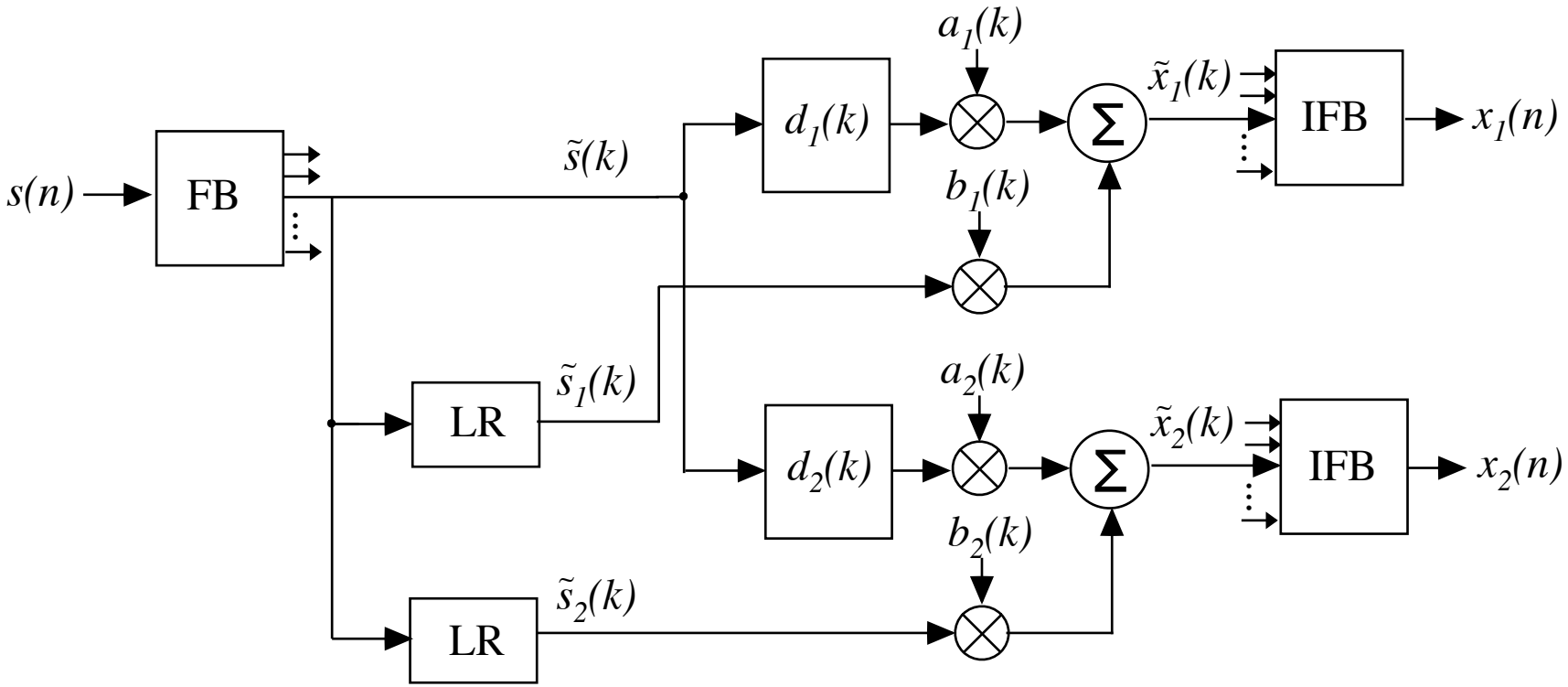


Numerical example:



Binaural Cue Coding (BCC)

Alternative ICTD/ICLD/ICC synthesis

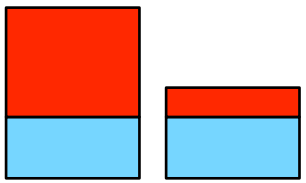


---

a_i, b_i : scale factors

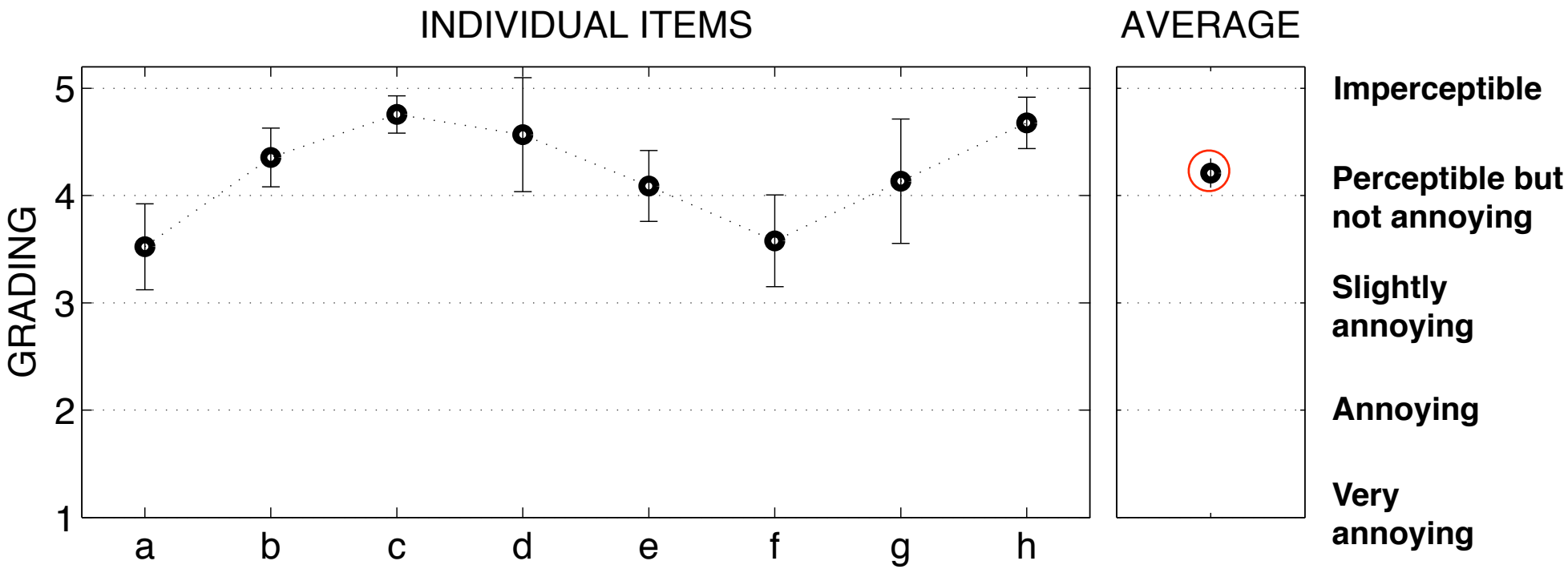
LR: late reverberation filter

Coherent/**incoherent** channel power:



Binaural Cue Coding (BCC)

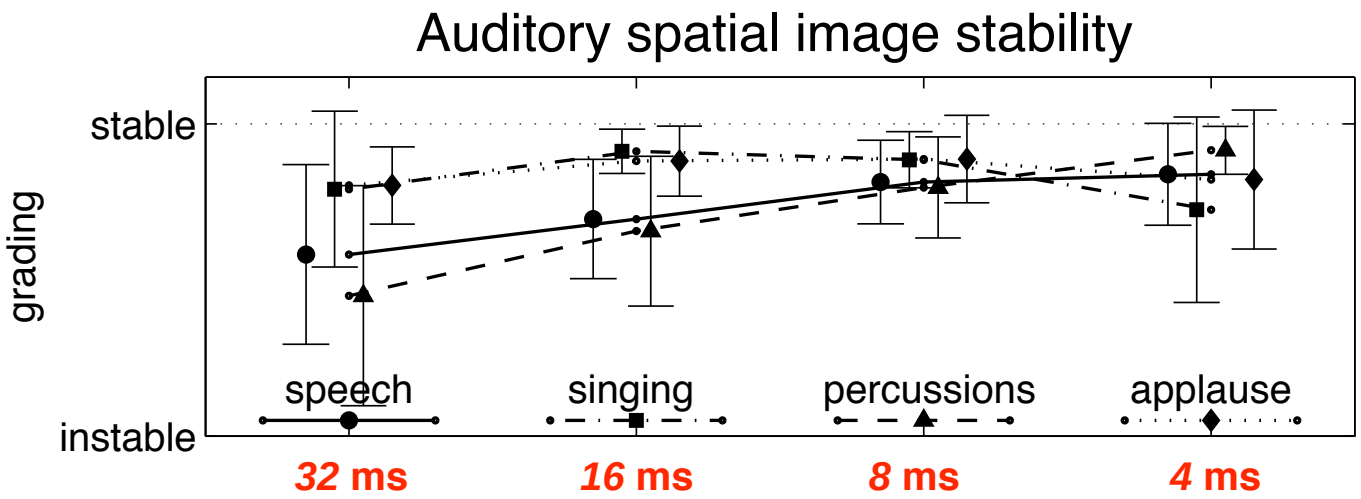
Subjective evaluation: 5-channel audio quality

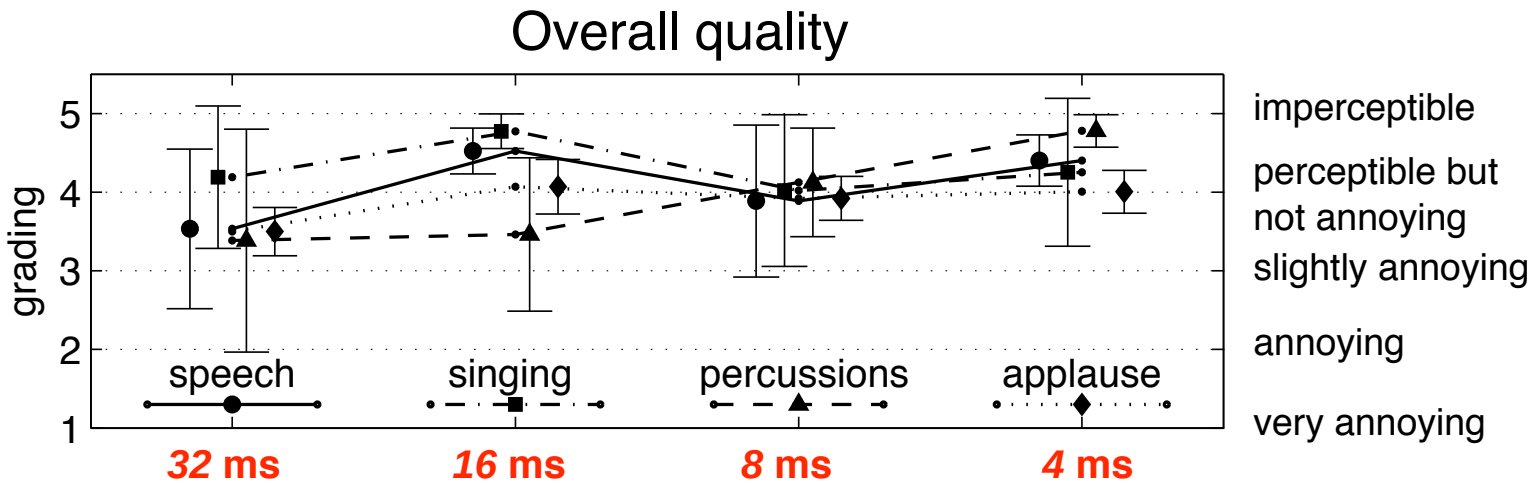


Hidden reference test (ITU-R BS.1116), loudspeaker listening, 9 subjects

Binaural Cue Coding (BCC)

Subjective evaluation: ICLD time resolution

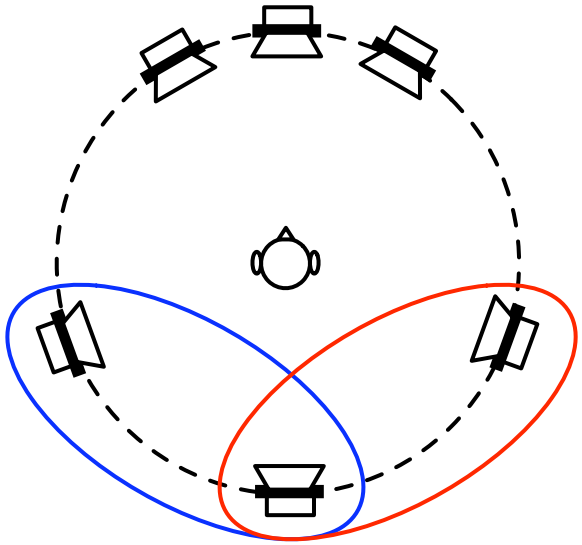




Variations of BCC

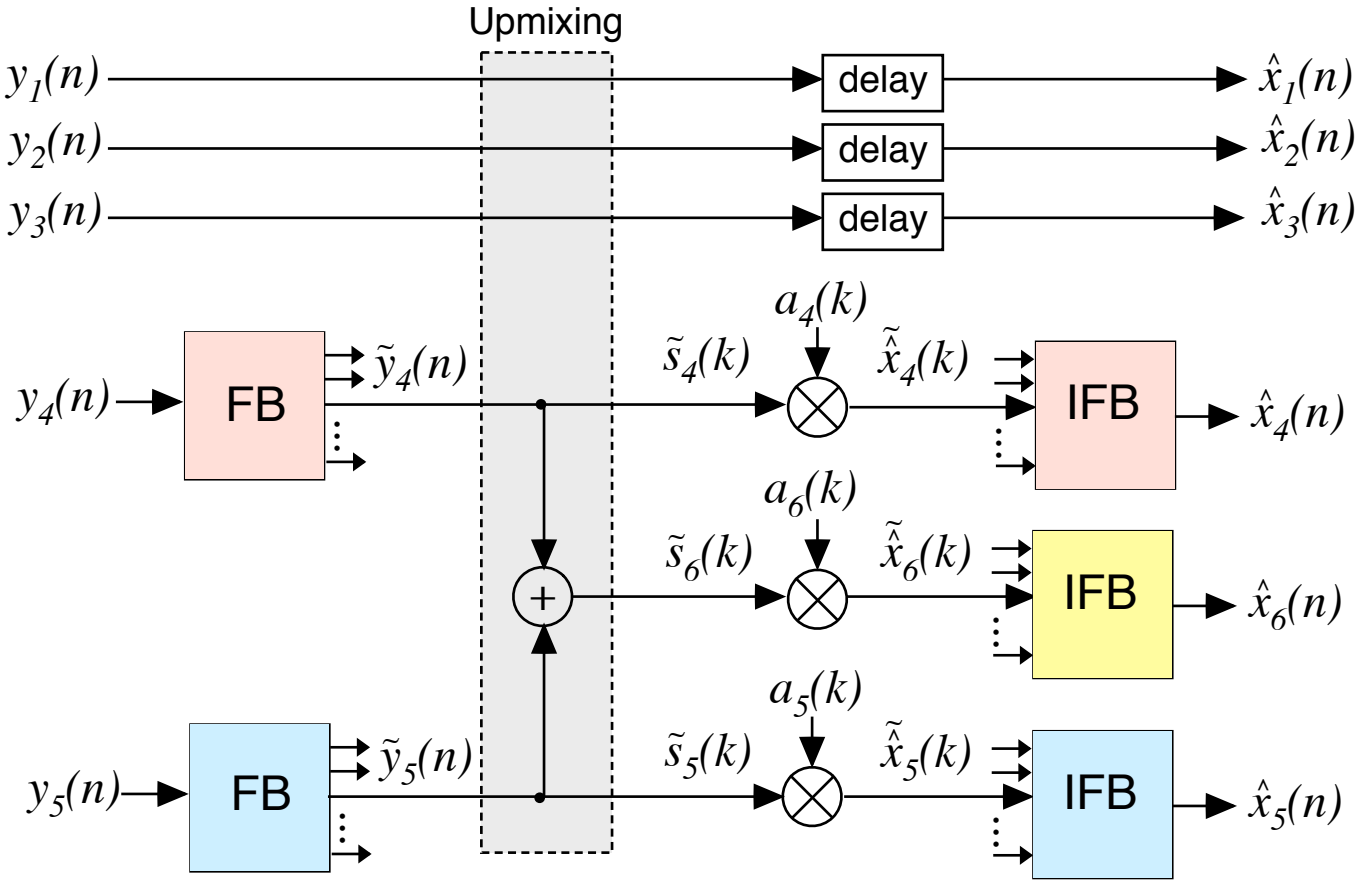
6-to-5 BCC: 5.1 backwards compatible coding of 6.1

Downmix:



Dolby Surround EX
loudspeaker positioning

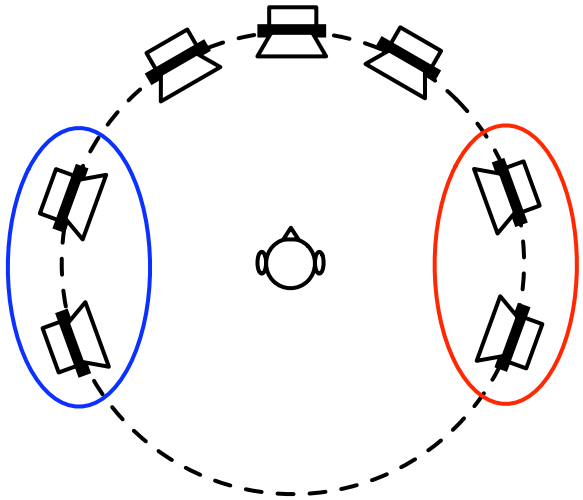
5-to-6 Synthesis:



Variations of BCC

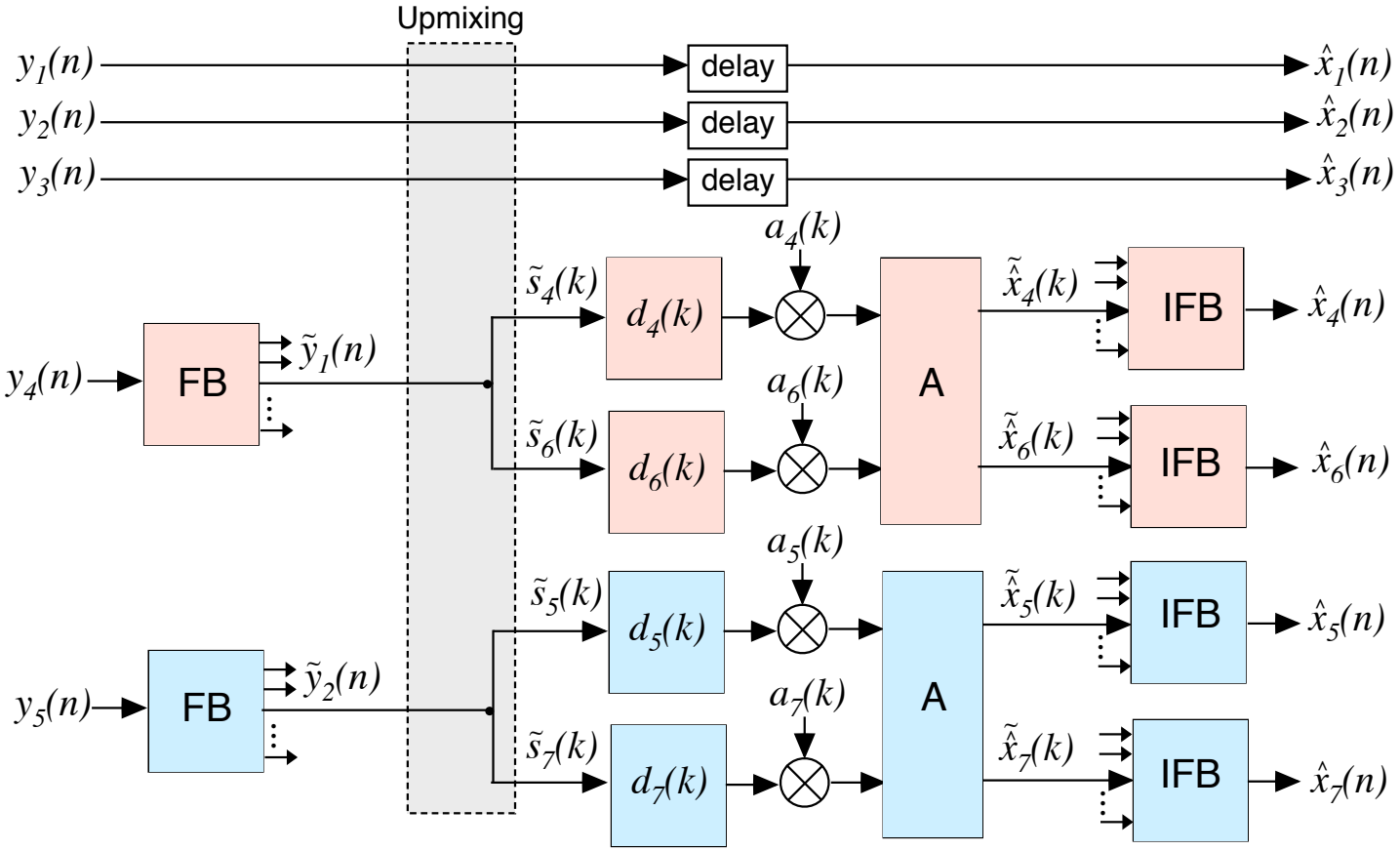
7-to-5 BCC: 5.1 backwards compatible coding of 7.1

Downmix:



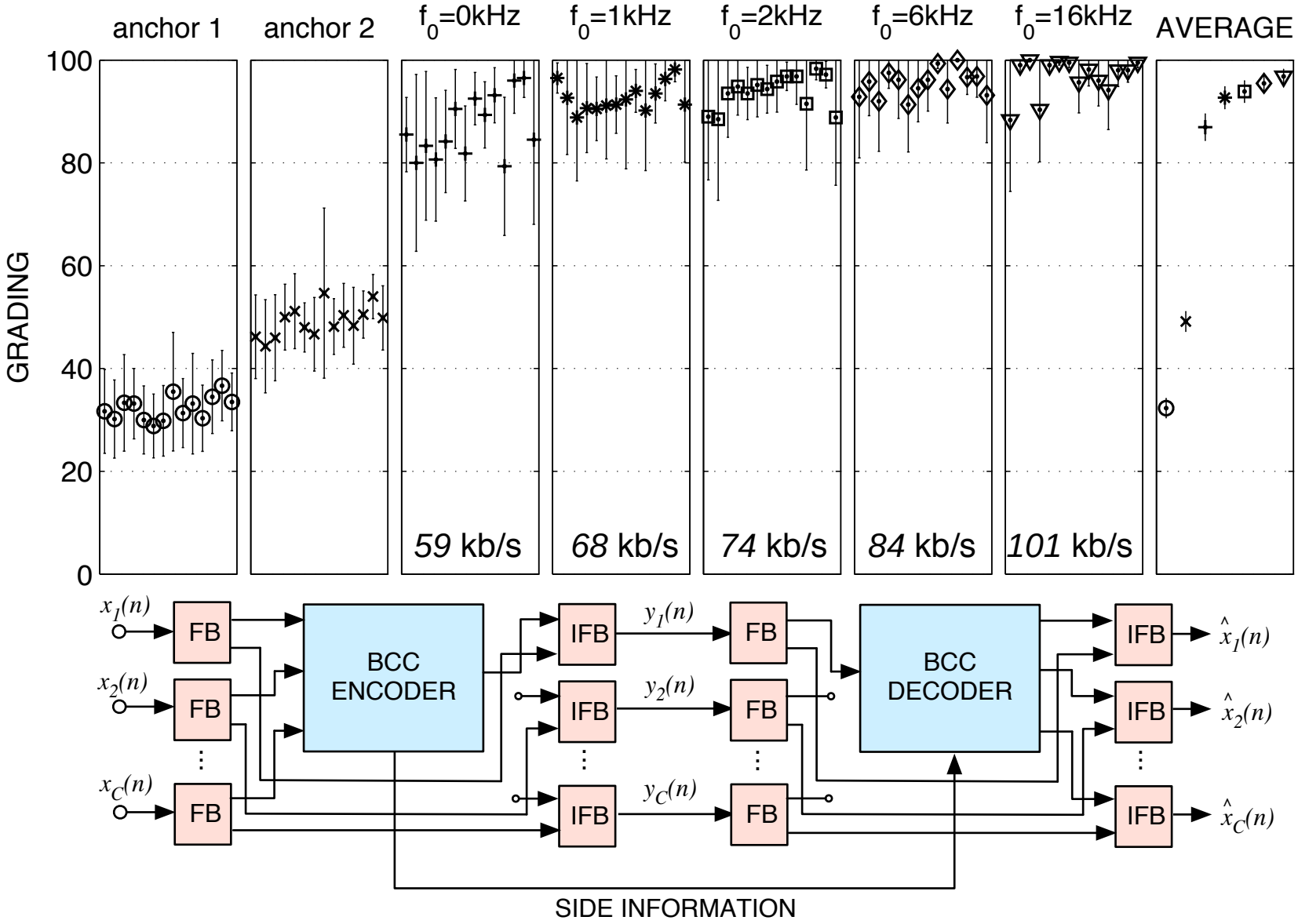
Lexicon Logic 7
loudspeaker positioning

5-to-7 Synthesis:



Variations of BCC

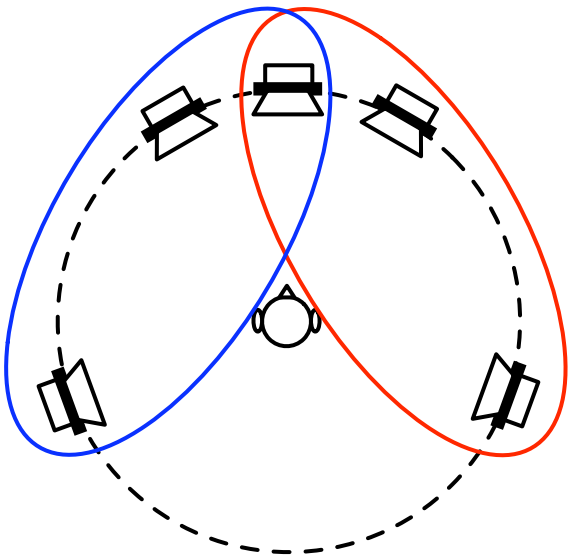
Hybrid BCC



Variations of BCC

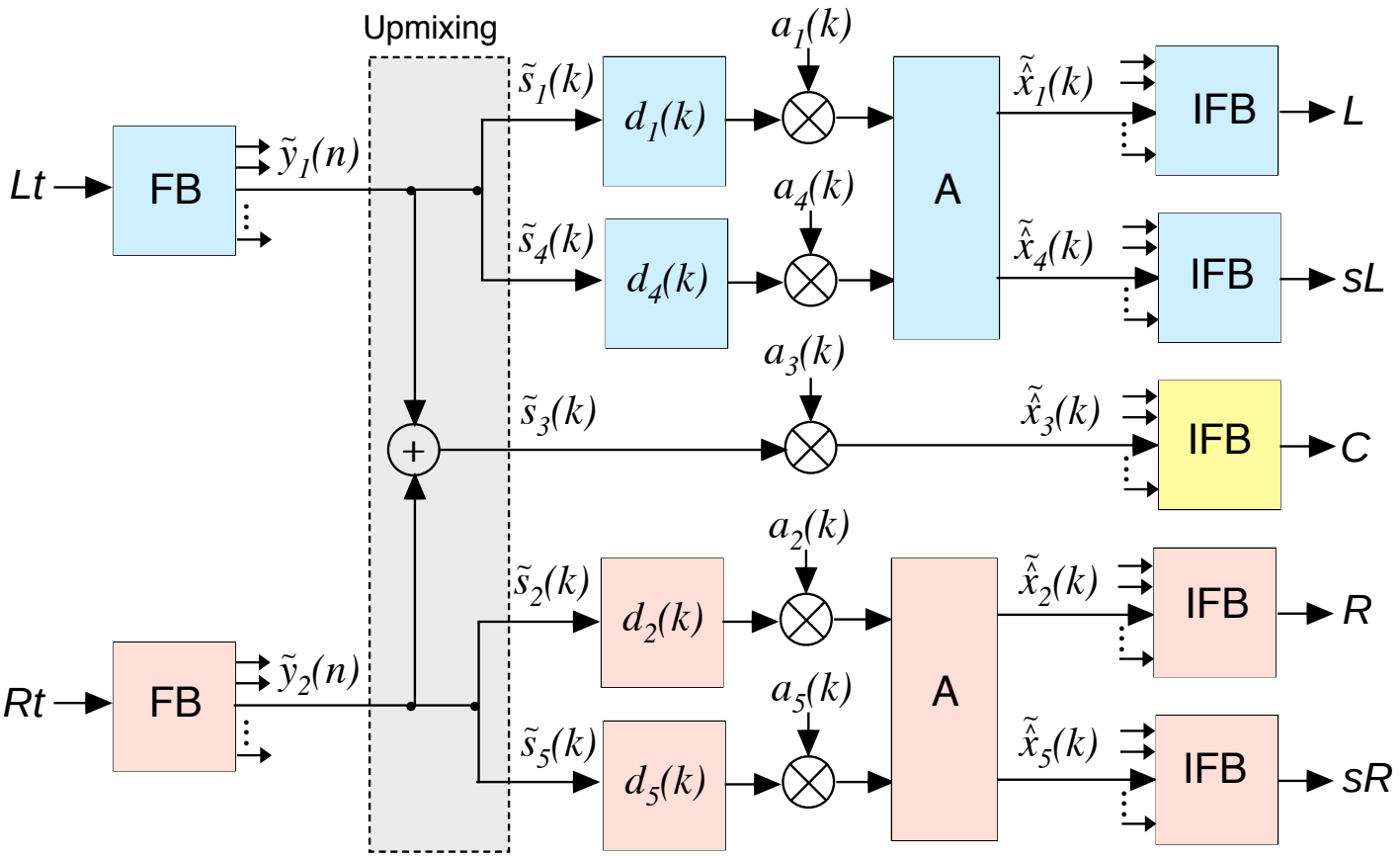
5-to-2 BCC: Stereo backwards compatible coding of 5.1

Stereo-downmix:



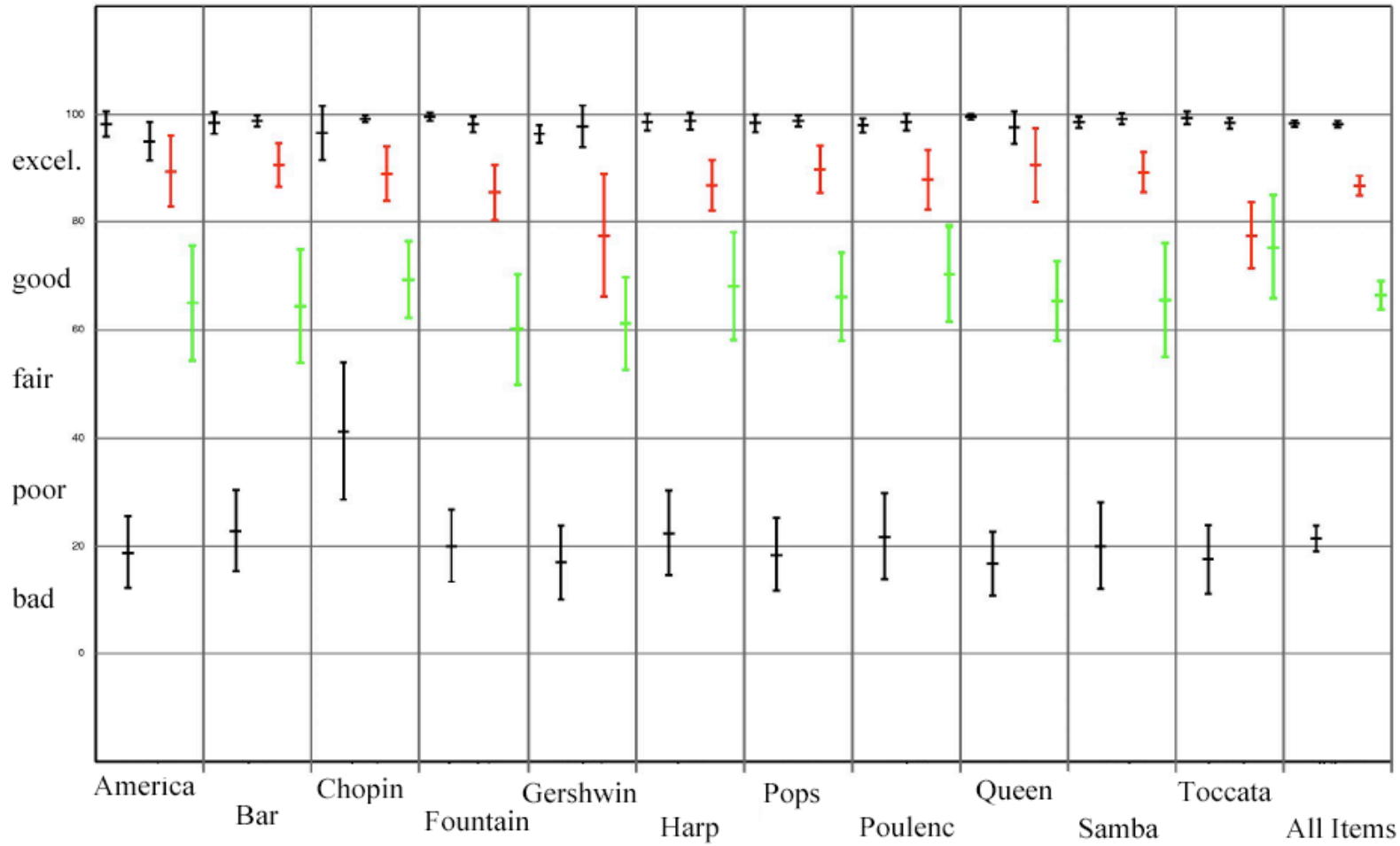
Lt = L + 0.7C + sL
Rt = R + 0.7C + sR

2-to-5 Synthesis:



Variations of BCC

Low complexity **5-to-2 BCC:** Subjective evaluation:
MUSHRA (ITU-R BS.1534)



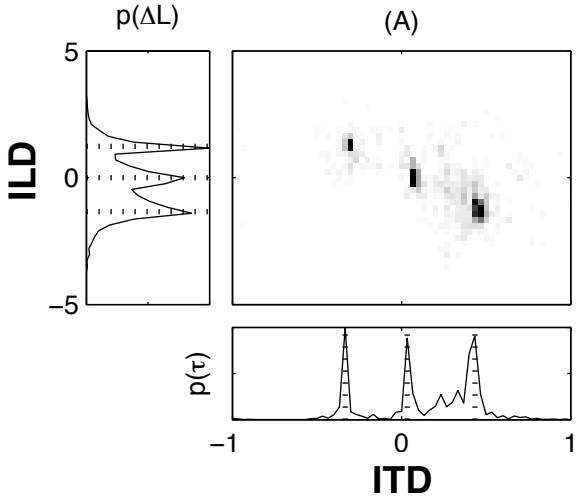
Hidden reference, Anchor, AAC, 5-to-2 BCC + MP3, Dolby Prologic II

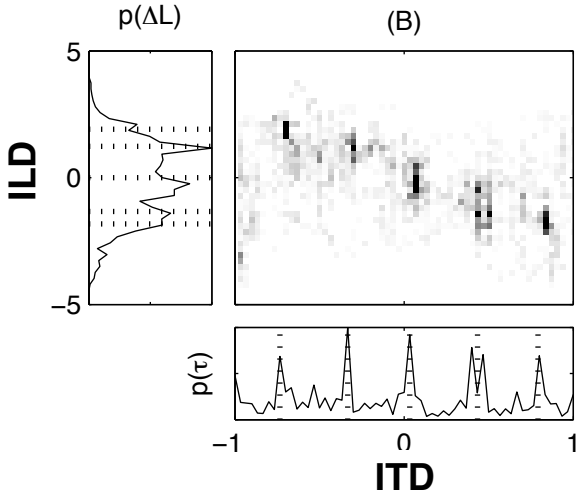
256 kb/s **192 kb/s** **PCM**

Source Localization in Complex Listening Scenarios

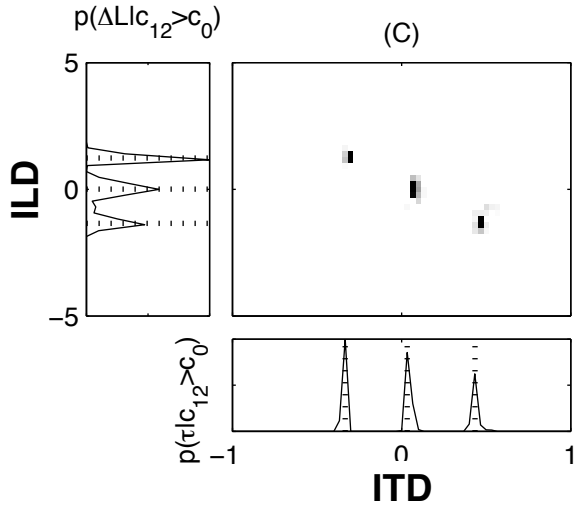
Three and five concurrent male speech sources:
(-30°, 0°, 30°, -80°, 80°, free-field)

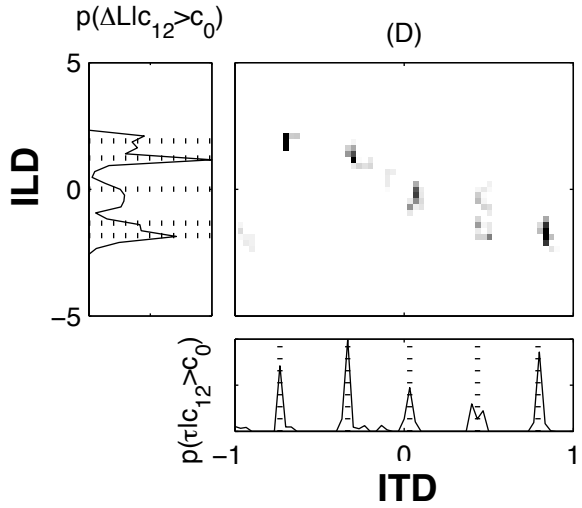
Without cue selection





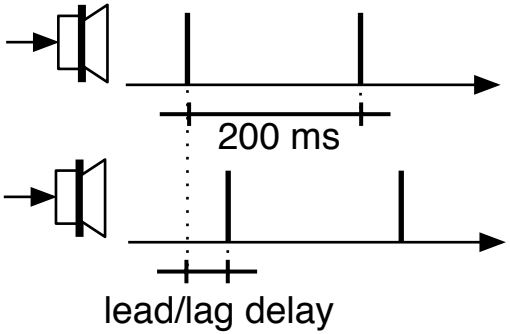
With cue selection

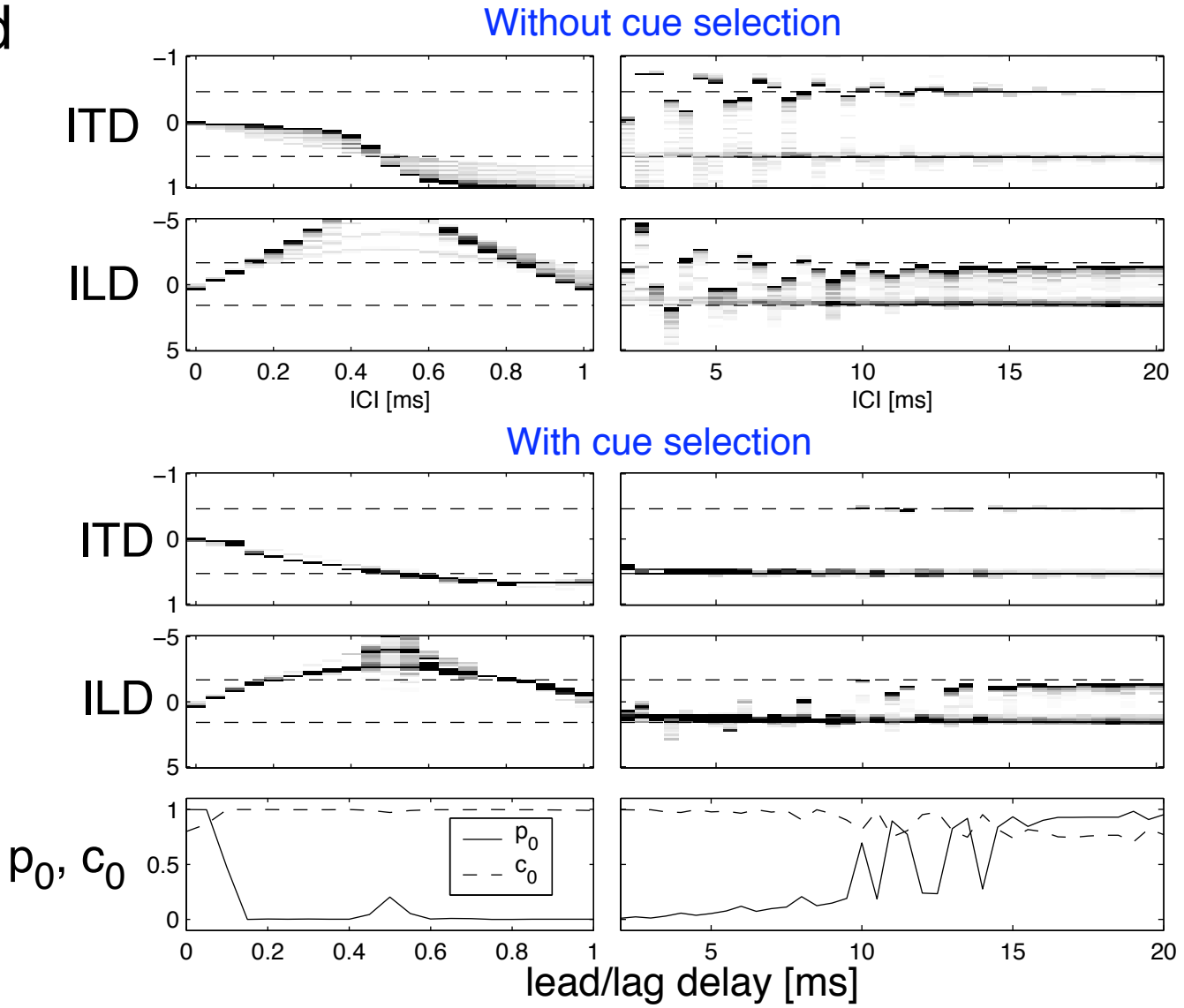




Source Localization in Complex Listening Scenarios

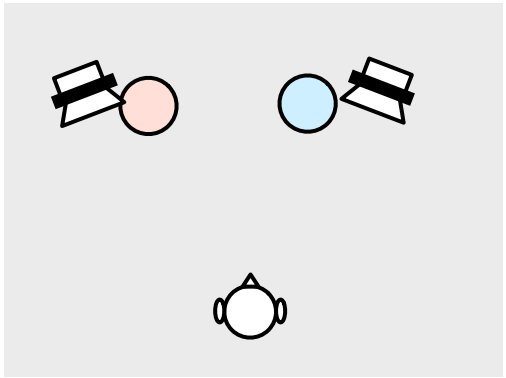
Cue selection: Three phases of the precedence effect, 500Hz critical band

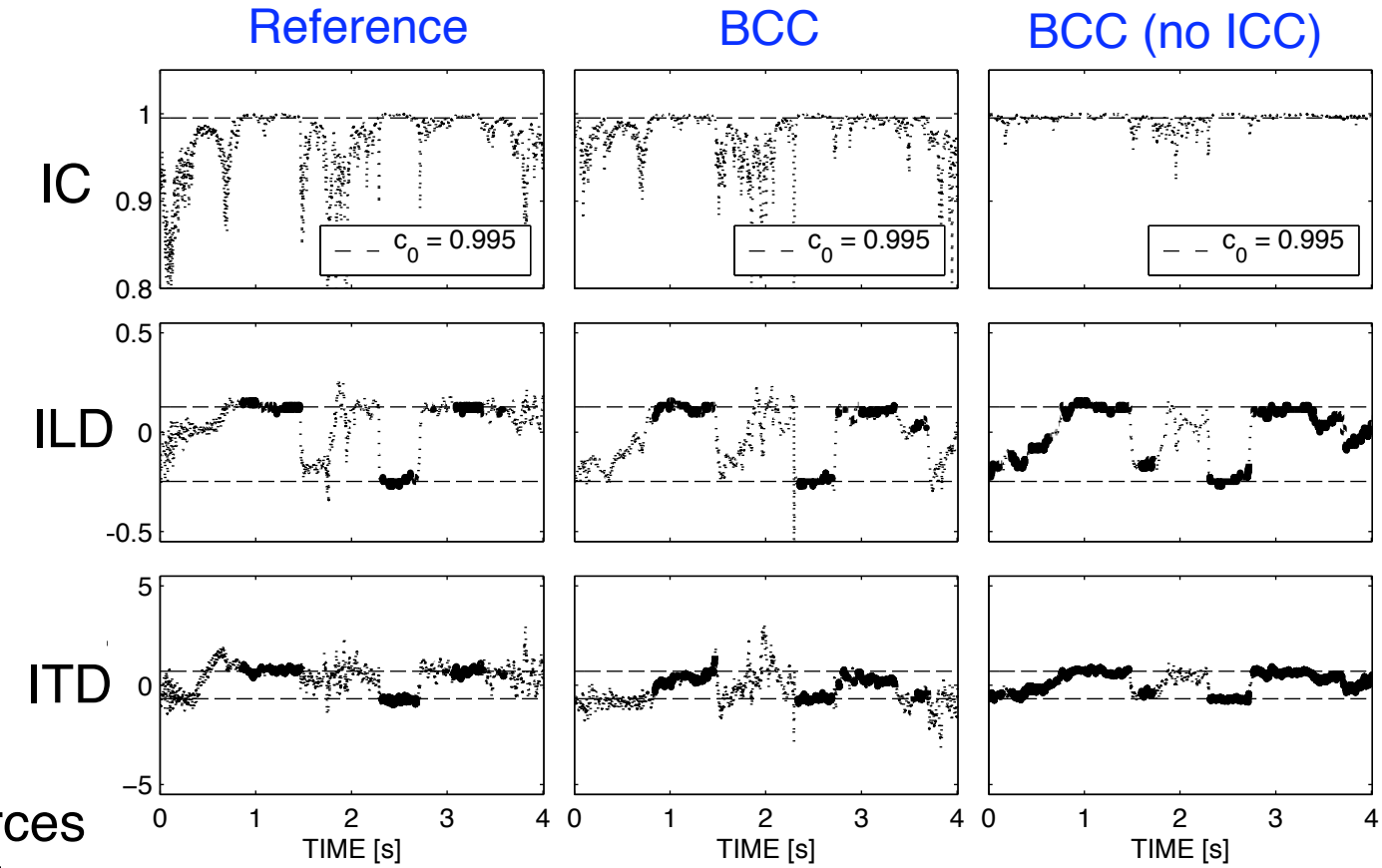




Source Localization in Complex Listening Scenarios

Amplitude panning, free-field, 500Hz critical band





Standard stereo setup
Two male speech sources
+-8dB amplitude panning.